

基于两阶段分层 DQN 的电动汽车实时充电调度策略

郑逸凡, 曾益, 易心雨, 李江晟, 余春晓, 王华文

(华东交通大学交通运输工程学院, 江西 南昌 330013)

摘要: 针对电动汽车充电路径规划中现有方法忽略排队随机性且难以兼顾实时性与全局最优的问题, 提出一种基于广义总成本 (TGC) 与分层深度 Q 网络 (DQN) 的路径优化方法。首先, 引入 M/M/c 排队模型量化充电等待时间的随机性, 通过 TGC 体系协助深度 Q 网络将非线性排队、行驶时间与分时电价统一为单一标量奖励。其次, 结合 Top-M 预筛选机制与动作有效性掩码来解决大规模动作空间导致的收敛难题, 并利用经验回放打破数据相关性。实验在构建的高保真仿真环境中进行性能评估, 在 Top-M 参数设为 10 时, 该策略的全局最优解命中率达 89.5%, 平均相对后悔值收敛至 0.0122, 单次推理耗时缩短至 0.35ms。

关键词: 电动汽车; 充电路径规划; 深度强化学习; 智能交通系统

中图分类号: TP 18

文献标志码: A

Real-Time Charging Scheduling Strategy for Electric Vehicles Based on a Two-Stage Hierarchical DQN

Zheng Yifan, Zeng Yi, Yi Xinyu, Li Jiangsheng, Yu Chunxiao, Wang Huawen

(School of Transportation Engineering, East China Jiaotong University, Nanchang 330013, China)

Abstract: To address the issues in existing electric vehicle charging path planning methods, namely the neglect of queuing randomness and the difficulty in balancing real-time performance with global optimality, this paper proposes a path optimization method based on Generalized Total Cost (TGC) and a hierarchical Deep Q-Network (DQN). First, an M/M/c queuing model is introduced to quantify the randomness of charging waiting time. The TGC framework assists the Deep Q-Network in unifying nonlinear queuing time, travel time, and time-of-use electricity price into a single scalar reward. Second, a Top-M pre-screening mechanism combined with an action validity mask is employed to resolve convergence difficulties caused by large action spaces, and experience replay is utilized to break data correlations. The method's performance is evaluated in a constructed high-fidelity simulation environment. When the Top-M parameter is set to 10, the proposed strategy achieves a global optimal solution hit rate of 89.5%, an average relative regret value converging to 0.0122, and a single inference time reduced to 0.35 ms.

Key words: Electric vehicles; Charging path planning; Deep reinforcement learning; Intelligent transportation systems

近年来, 为响应国家绿色经济发展, 电动汽车行业发展迅速。根据公安部数据显示, 截至 2025 年 6 月底, 新能源汽车保有量增至 3689 万辆 (纯电动 2553.9 万辆)^[1]。随着我国新能源汽车保有量的增长, 如何精准规划电动汽车充电路径、协调电动汽车与充/换电设施之间的三维关系成为了目前亟待解决的问题。对于电动汽车充电路径规划的研究, 现有的方法大多是基于全局搜索方法, 但随着城市路网的不断发展, 此

收稿日期: 2026-05-11

基金项目: 国家自然科学基金项目 (72161011)

方法往往难以实现广义总成本的最小化目标；同时，从现实情况出发，分时电价与到达流叠加会导致峰段排队现象，旧有基于历史数据取平均的计算充电等待时间的方法可能会对排队过程的随机性考虑不足。

许多学者对电动汽车充电路径优化问题进行了广泛研究，主要集中于以下两个方面。a)围绕带时间窗的电动汽车路径优化，将行驶路径、充电地点与运营成本统一到路径-充电一体化的优化框架中。Wang 等通过高斯混合聚类降低问题规模、再用改进多目标遗传算法与禁忌搜索的杂交算法优化路径与充电站选择，算例表明该方法在降低成本、提高效率方面优于多种经典多目标算法^[2]。Catay 和 Sadati 则将电动车路径问题推广为带时间窗与移动充电站的 EVRPTW-MCS，提出结合变邻域搜索与精确算法的 matheuristic 框架^[3]。Liu 将 EV 充电调度（EVCS）形式化为分层混合变量优化问题，将站点选择、充电选项与充电量三类决策统一建模^[4]。b)从充电调度与负荷管理出发，通过能量管理、数据驱动建模或强化学习等方法间接实现能耗与经济成本的最小化。Yao 等针对停车场场景，将实时充电与需求响应（DR）结合，将充电调度建模为二元优化问题，通过凸松弛替代穷举搜索求解近似最优解^[5]。Lu 等通过蒙特卡洛模拟刻画 EV 充放电的不确定性，显著降低了社区整体能耗与用户成本^[6]。Wang 等提出一种数据驱动的 EV 聚合负荷建模方法，构建刻画充电需求柔性的数学模型^[7]。在强化学习层面，Tuchnitz 等提出基于强化学习的车队充电协调系统，采用神经网络作为函数逼近器，为任意规模的 EV 车队生成前瞻性的协调充电计划^[8]。Park 和 Moon 提出集中训练-分散执行的多智能体深度强化学习方法^[9]。

综上所述，已有研究大多集中于对站点选择或电动汽车车辆调度进行单独优化。同时，大部分研究并未充分考虑充电等待时间时可能导致的排队问题。另外，传统的全局搜索方法（如蚁群算法、遗传算法）在处理此类大规模组合优化问题时，通常需要多次迭代才能收敛，且算法高度依赖具体路网结构，这种方法或将难以实现广义总成本的最小化目标，也无法适应充电基础设施日益密集的现实发展趋势。针对这些问题，本文主要工作如下：

1) 基于 TGC 框架构建电动汽车充电路径优化模型，将随机过程中的非线性排队时间、确定性的行驶时间与动态电价统一映射为单一维度的广义成本。

2) 采用 M/M/c 排队模型来模拟充电站的排队行为，通过估算充电站的平均排队时间来衡量充电站的服务能力与负载情况，以提升用户在路径选择时的效率。

3) 构建多目标优化的强化学习环境来实现电动汽车充电路径的智能规划；结合 Top-M 预筛选机制与动作有效性掩码相结合的分层深度 Q 网络(DQN)架构，解决大规模动作空间导致的收敛难题。

1 模型构建

1.1 模型假设

为验证模型有效性，提出如下假设：充电站的服务压力分布与路网的接近中心性及 POI 需求强度呈现强正相关；在两阶段决策框架中，存在一个较小的候选集大小 K ，使得 Top- K 策略的收益收敛于全局最优解；广义总成本（TGC）对时间价值参数 β 的一阶导数 $d(T)/d\beta$ ，在数值上近似等于用户的总耗时；即使站点处于高服务强度状态（ $\rho \rightarrow 0.95$ ），系统的尾部延迟仍表现为数值有界且收敛；智能体能够通过与环境的交互学习，使其决策策略的相对后悔值收敛至零附近的极小邻域。

1.2 TGC 定义与构成

本文定义广义总成本 C_{TGC} 为用户完成一次充电决策所付出的总经济成本与时间成本之和，其数学表达

式如下：

$$C_{TGC} = C_f + \beta \cdot (T_d + T_w + T_c) \quad (1)$$

其中， C_f 为用户支付的实际充电费用（元）； T_d 为车辆从当前位置沿路网行驶至目标充电站的预估耗时（h）； T_w 指车辆到达站点后，因充电桩满员而产生的排队等待时间（h）； T_c 指车辆从当前剩余电量 S 充至目标电量所需的物理时间（h）， S 为荷电状态； β 为时间价值系数（元/h）。

在量纲换算上，充电费用 C_f 的计算公式为：

$$C_f = \pi_j E_{cap} (S_{tar} - S_i) \quad (2)$$

其中 π_j 为第 j 种充电场景下的电价，单位为元/kWh； E_{cap} 为电池的总容量，单位为 kWh； S_{tar} 为充电结束时期望达到的电池荷电状态，取值范围为[0,1]； S_i 为充电开始前的电池荷电状态，取值范围为[0,1]；时间项 T_d 、 T_w 、 T_c 均以小时计，乘以 β 后折算为货币成本。

在广义总成本的构成体系中，时间成本由车辆行驶时间、充电时间、等待时间所组成。车辆行驶时间计算公式如下：

$$T_d(u, j) = \frac{D_{net}(u, j)}{V_{avg}} \times 60 \quad (3)$$

其中， $D_{net}(u, j)$ 为节点 u 到节点 j 的最短路径距离， V_{avg} 为城市路网的平均通行速度。

充电时间的计算公式为：

$$T_c = \frac{(1-S) \times E_{cap}}{P} \times 60 \quad (4)$$

其中， P 为目标站点的充电桩输出功率（kW）。

1.3 M/M/c 队列模型

等待时间 T_w 主要源于多用户对有限充电桩的竞争，具有高度的随机性。为准确量化这一随机成本，本文将城市路网中的每一个充电站点 j 建模为一个独立的 M/M/c 排队系统。为量化系统的负载水平，定义交通强度 ρ 来描述站点的服务饱和度：

$$\rho_j = \frac{\lambda_j}{c_j \mu} \quad (5)$$

其中， λ_j 为每小时到达的车辆总需求， $c_j \mu$ 为该站点所有充电桩每小时提供的最大服务总量， ρ_j 为判定系统稳定性的指标。

为计算车辆的期望等待时间，首先推导系统处于各种状态的稳定概率，当充电站内没有任何车辆时，

概率 P_0 可表示为:

$$P_0 = \left[\sum_{k=0}^{c_j-1} \frac{(c_j \rho_j)^k}{k!} + \frac{(c_j \rho_j)^{c_j}}{c_j! (1 - \rho_j)} \right]^{-1} \quad (6)$$

当充电站内所有充电桩均被占用时, 车辆必须进入排队序列, 引入延迟概率 P_w :

$$P_w = \frac{(c_j \rho_j)^{c_j}}{c_j! (1 - \rho_j)} \cdot P_0 \quad (7)$$

综上, 车辆在队列中的平均等待期望时间 T_w 最终可推导为:

$$T_w = \frac{P_w}{c_j \mu (1 - \rho_j)} \quad (8)$$

1.4 问题建模与状态空间

在深度强化学习的框架下, 充电路径优化问题可表述为一个马尔科夫决策过程, 由元组 $\langle S, A, R, \gamma \rangle$ 构成。其中, S 为环境状态, A 为离散动作空间, R 为全局累积奖励, γ 为折扣因子。

对于任意决策步 t , 状态向量 S_t 由车辆物理状态与候选集环境状态组成, 其数学构造形式如下:

$$S_t = [s_v \oplus s_1 \oplus s_2 \oplus \dots \oplus s_M] \quad (9)$$

其中, $\oplus$ 表示向量拼接操作, M 为预筛机制设定的候选数量, s_v 为当前剩余电量, 取值范围为 $[0,1]$ 。

动作空间 A 被定义为离散的有限集合 $\{A\} = \{0, 1, 2, \dots, M-1\}$, 动作 $a_t = k$ 代表智能体选择 Top-M 候选列表中的第 k 个站点作为目标充电站。

1.5 奖励函数设计

强化学习的优化目标是最大期望累积奖励, 而驾驶员的实际目标是最小化服务成本, 为统一二者关系, 本文将奖励函数定义为动作 a_t 对应的负值广义总成本, 即:

$$r_t(s_t, a_t) = -T(s_t, a_t) \quad (10)$$

其中, $-T(s_t, a_t)$ 为负值广义总成本, s_t 为智能体在时刻 t 观察到的环境信息。

展开广义总成本构成项, 完整的奖励函数表示为:

$$r_t(s_t, a_t) = -[C_f(a_t) + \beta \cdot (T_d(s_t, a_t) + T_w(a_t) + T_c(s_t, a_t))] \quad (11)$$

其中, $C_f(a_t)$ 为电费成本, $T_d(s_t, a_t)$ 为行驶时间, $T_w(a_t)$ 为等待时间, $T_c(s_t, a_t)$ 为充电时间, $-\beta T_d$ 为

行驶惩罚， $-\beta T_w$ 具体含义为拥堵惩罚， $-C_f$ 为价格惩罚。

1.6 DQN 训练机制

为捕获路径规划中复杂的非线性关系，本文设计了包含两个全连接层的隐藏结构，每层神经元数量设定为 256，层与层之间采用 ReLU 作为激活函数。网络的输出层采用线性激活函数，包含 M 个神经元。

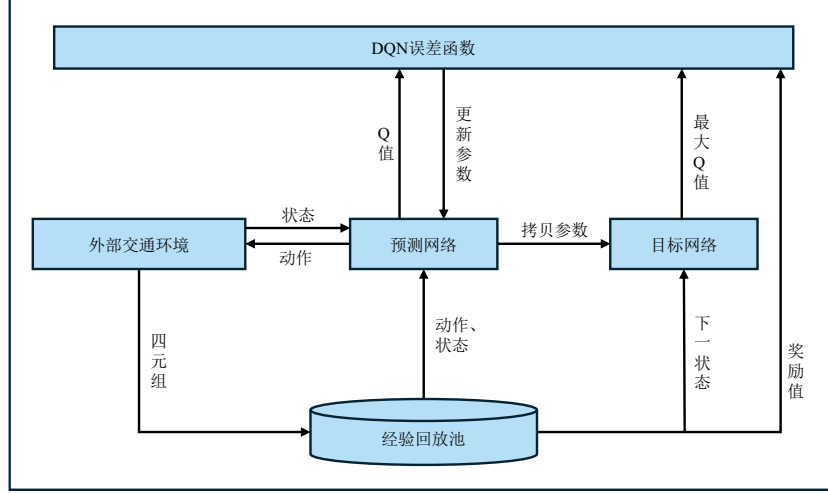


图 1 DQN 算法流程图

Fig.1 Flowchart of the DQN algorithm

如图 1 所示，本文集成了经验回放与目标网络两种机制。

经验回放机制是指，在仿真交互过程中，智能体产生的每一条轨迹数据并不直接用于训练，而是被储存在一个固定容量的经验回放池中。目标网络则主要用于计算时序差分目标值，公式如下：

$$y_i = r_i + \gamma \cdot \max_{a'} Q(s_{i+1}, a'; \theta^-) \quad (12)$$

其中， r_i 为即时奖励； $\max_{a'} Q(s_{i+1}, a'; \theta^-)$ 为目标网络的最大 Q 值； $Q(s_{i+1}, a'; \theta^-)$ 为目标网络的动作价值函数。

由于 DQN 训练本质上是一个回归问题，本文采用均方误差作为损失函数，具体计算公式如下：

$$L(\theta) = E_{(s,a,r,s') \sim D} \left[\left(r + \gamma \max_{a'} Q(s', a'; \theta) - Q(s, a; \theta) \right)^2 \right] \quad (13)$$

其中， $E_{(s,a,r,s') \sim D}$ 为期望算子， $Q(s', a'; \theta)$ 为目标网络输出的动作价值。

为了处理充电站推荐场景中稀疏且非平稳的梯度信号，最小化该损失函数，本文采用了 Adam 优化器。训练统一采用 PyTorch/Stable Baselines 3 实现，学习率设为 $1e-4$ ，批次大小 64，折扣因子 0.99，经验回放池容量 100000，探索率从 1.0 线性衰减至 0.05；收敛判据为滑动平均奖励曲线进入平台期且验证集平均相对后悔值稳定在 0.02 以内，最终模型在 100000 步训练后用于测试。

综上，本文的 DQN 决策模型训练流程分为以下步骤：解析南昌市交通路网拓扑数据与充电站分布元数据，构建基于 Gym 框架的强化学习仿真环境；将状态向量输入 DQN 网络；计算该动作产生的广义总成本并转化为奖励信号，将该次交互元组存入经验回放池后；监控经验回放池的数据量，当样本数量达到预设阈值后，随机采样小批量数据用于网络训练；执行基于梯度下降的神经网络参数更新，每间隔 k 步将训

练网络的参数同步至目标网络，直至模型收敛，完成训练。

2 实验分析

2.1 实验说明与基本参数设定

本文选取江西省南昌市为研究区域，以 OSM 南昌市中心城区路网数据为基础数据集。通过聚合餐饮、住宅、商业等 POI 数据，采用自适应半径搜索与图核平滑算法，得到连续的充电需求强度场如图 2 所示，图 2 中横轴标注的“Lon”对应经度，纵轴标注的“Lat”对应纬度（单位：°），色标标注的“value (0-1)”表示归一化值（无量纲）。由于本文后续未使用经度、纬度及归一化值参与公式运算，此处不引入单字母变量符号，图中标签仅为坐标轴与色标的说明文字。具体实现方法如下。

实验路网通过 OSMnx 获取南昌市中心城区道路网络，主体路网包含 38133 个节点和 56167 条有向边；充电站样本为 17 个公共站点。训练与测试任务均以普通路网节点作为起终点，剔除充电站节点、孤立节点和不可达节点；每个任务随机抽取 O-D 对，初始 S 服从 $[0.30, 0.80]$ 区间的均匀分布，目标 S 设为 0.80，并通过能量可达性掩码剔除无法到达的候选站点。测试集含 200 个独立随机任务，与训练交互样本相互独立。

引入“全局覆盖率”作为尺度标定的约束指标，设路网节点集合为 V ，POI 结合为 P ，给定半径 r 时，节点 v 的 POI 计数为：

$$c_v(r) = \left| \left\{ p \in P \mid d_{hav}(v, p) \leq r \right\} \right| \quad (14)$$

其中， d_{hav} 为基于地球球面的 Haversine 距离。在此基础上，对所有节点的覆盖率为非零计数节点占总节点数的比例为：

$$\text{cov}(r) = \frac{1}{|V|} \sum_{v \in V} \mathbb{I}\{c_v(r) > 0\} \quad (15)$$

其中， $\mathbb{I}$ 为指示函数，为消除绝对计数的量纲影响并保证数值稳定性，再进一步对 $c_v(r^*)$ 进行 Min-Max 归一化，得到未平滑的节点级需求强度：

$$D_{raw}(v) = \frac{c_v(r^*) - \min(C) + \varepsilon}{\max(C) - \min(C) + \varepsilon} \quad (16)$$

其中， $c_v(r^*)$ 表示在选定的半径 r^* 范围内，节点 v 周围所覆盖的 POI 的数量； $\max(C)$ 和 $\min(C)$ 分别表示所有节点中 POI 计数的最大值与最小值； ε 为极小正数。本文未对孤立节点进行额外的阈值截断，旨在最大限度保留原始数据的分布特征。

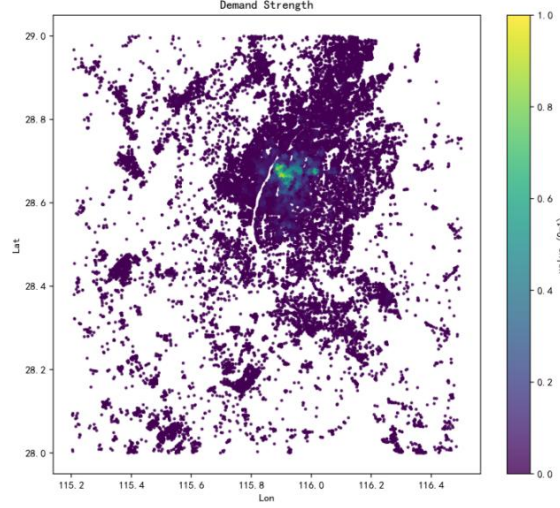


图 2 未平滑需求强度在研究区路网的空间分布

Fig.2 Spatial distribution of unsmoothed demand intensity in the road network of the study area

为了在路网尺度上获得空间连贯、噪声可控的需求强度指标，本文引入基于路网拓扑的图核平滑（Graph Kernel Smoothing）。设路网为无向图 $G=(V,E)$ ，其中， V 为节点集合， E 为边集合，令 $x^{(0)}$ 为未平滑需求强度向量，一阶邻域加权平均的线性图平滑模型为：

$$x_v^{(t+1)} = \alpha x_v^{(t)} + (1-\alpha) \cdot \frac{1}{|N(v)|} \sum_{u \in N(v)} x_u^{(t)}, v \in V \quad (17)$$

其中 $N(v) = \{u \in V : (v,u) \in E\}$ 为节点 v 的一阶邻接集合； $\alpha \in [0,1]$ 为平滑衰减系数；若 $N(v) = 0$ （孤立节点），则令 $x_v^{(t+1)} = x_v^{(t)}$ 。上述可写为矩阵形式：

$$X^{(t+1)} = \alpha X^{(t)} + (1-\alpha) D^{-1} A X^{(t)} \quad (18)$$

其中， $A \in R^{|V| \times |V|}$ 为无向邻接矩阵， D 为度矩阵， $D^{-1}A$ 是按行归一的随机游走算子。取平滑步长 $\alpha=0.7$ 。

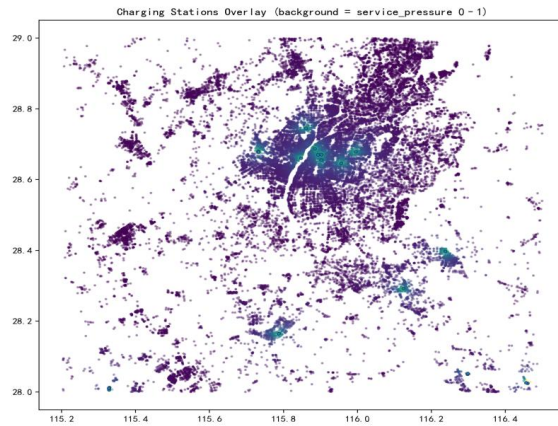


图 3 充电站分布图

Fig.3 Distribution Map of charging stations

本文将充电站点划分为三个电价档位，具体空间分布与服务压力图如图 3 所示，表 1 为不同档位站点的分布特征与服务能力，其余模型参数见表 2、3。表 3 中 $\rho_{\max}$ 为目标荷电状态上限， π 为分时电价（元）， T_{cov} 为需求强度覆盖半径（m）。

表 1 充电站点属性统计与分档特征
Table 1 Statistics and Tiering Characteristics of Charging Station Site Attributes

电价档位	站点数量	平均服务台数	地理特征描述	典型区域
1.08 元/kWh	6	5.50	低需求密度，交通末梢	城市边缘/居住区
1.25 元/kWh	5	5.40	中等需求，次干道	一般商业区/开发区
1.58 元/kWh	6	6.17	高需求密度，高中心性	市中心核心商圈
总计/均值	17	5.71	-	-

表 2 DQN 算法超参数清单
Table 2 List of Hyperparameters for the DQN Algorithm

参数	总训练步数	经验回放池	批次大小	折扣因子	探索率
取值	100000	100000	64	0.99	1.0→0.05

表 3 仿真环境参数清单
Table 3 List of Simulation Environment Parameters

参数	P /kW	E_{cap} /kWh	$\rho_{\max}$	β /元/h	π /元	T_{cov} /m	M
取值	90	60	0.95	50	{1.08, 1.25, 1.58}	~850	10

2.2 实验结果与分析

2.2.1 评价指标

本文将广义总成本作为衡量方案优劣的核心指标，引入 TOP-1 命中率来衡量智能体是否成功命中全局最优解，其含义为智能体推荐的最优候选站点与全局真正最优站点一致的任务所占的百分比，公式如下：

$$A_{top} - 1 = \frac{1}{N} \sum_{i=1}^N (a_{agent}^i = a_{opt}^i) \times 100\% \quad (19)$$

其中， N 为测试集样本量， a_{agent}^i 为智能体选择的站点索引， a_{opt}^i 为理论最优索引。

引入相对后悔值(Relative Regret, R_{rr})作为回归指标，用于量化决策的偏离程度，公式如下：

$$R_{rr} = \frac{1}{N} \sum_{i=1}^N \frac{T_{agent}^i - T_{opt}^i}{T_{opt}^i} \times 100\% \quad (20)$$

其中， T_{agent}^i 为智能体决策产生的广义总成本， T_{opt}^i 为理论最优决策对应的广义总成本。

考虑仿真过程的随机性，本文使用包含 200 个独立随机任务的测试集，在确定性策略下进行全量评估，表 4 汇总了关键性能指标。

表 4 决策性能最终量化评估表 (N=200)

Table 4 Final Quantitative Evaluation Table of Decision Performance (N=200)				
评估指标	平均相对后悔	95%置信区间	TOP-1 命中率	平均决策耗时
数值结果	0.0122	[0.0060, 0.0183]	89.5%	0.35 ms

2.2.2 基于 DQN 的路径优化分析

本文选取固定测试集任务中的典型算例进行对比分析。边界条件设定如下：时间价值系数 $\beta=50$ 元/小时，初始剩余电量 $S=46.6\%$ ，算例选取 $\text{task_id}=96$ （起点节点 1219、终点节点 4684），在相同环境与参数条件下，对比与带时间窗的 Dijkstra 算法之间的差异与成本来源。

针对该算例，本文将起点附近的候选充电站标记为 $S_1 \sim S_4$ ，并计算各站点对应的广义总成本及其时间分量分解，结果如表 5 所示，表中 d_{to} 为驶向耗时， d_{from} 为驶离耗时。

表 5 典型算例对照表

Table 5 Comparison Table of Typical Test Cases								
站点	充电价格/元/kWh	服务台数/个	广义总成本/元	d_{to} /h	T_d /h	T_w /h	T_c /h	d_{from} /h
S_1	1.25	7	78.5210	0.5209	0.0109	0.0180	0.2225	0.8183
S_2	1.58	7	146.9556	1.2626	1.2626	0.0718	0.2225	0.7493
S_3	1.08	4	173.7827	0.9561	0.9561	0.2704	0.2225	1.5940
S_4	1.08	4	99.5452	0.5110	0.5110	0.0255	0.2225	0.7994

为了进一步揭示 DQN 智能体在经济成本与时间成本之间的博弈过程，图 4 对模型筛选出的前 5 个候选站点进行广义总成本的结构化拆解。

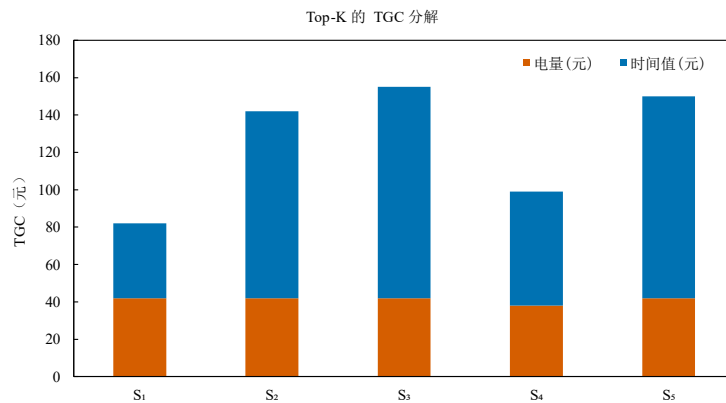


图 4 广义总成本拆解图

Fig.4 TGC Cost Breakdown Chart

通过对比最优解 S_1 与次优解 S_4 ，可以观察到 DQN 模型的决策特征： S_1 站点虽然并非最便宜的充电站点，但距离顺路且没有拥堵现象，综合表现最好；而 S_4 站点虽然充电价格最低，但时间成本过高，可能地理位置偏远或存在排队现象。智能体并未受到单纯追求低价的贪婪算法影响，而是捕捉到了 S_4 站点时间成

本损耗过大的负收益关系，从而剔除了 S_4 站点，选择了综合广义总成本更优的 S_1 。

在传统交通导航场景下，若用户仅依据静态网络信息选择充电站，可通过带时间窗的 Dijkstra 算法进行路径规划。本文将带时间窗 Dijkstra 基线定义为：对所有满足荷电状态可达性和时间窗约束的候选站点，分别求 $O \rightarrow CS$ 与 $CS \rightarrow D$ 的最短行驶时间路径，以总行驶时间最小作为站点选择规则；选定站点后仍按相同广义总成本公式重新计入电费、等待时间和充电时间，保证比较口径一致。本文调用已通过训练的 DQN 网络与带时间窗的 Dijkstra 算法在相同的实验环境下分别核算全局成本，对比图如图 5 所示。

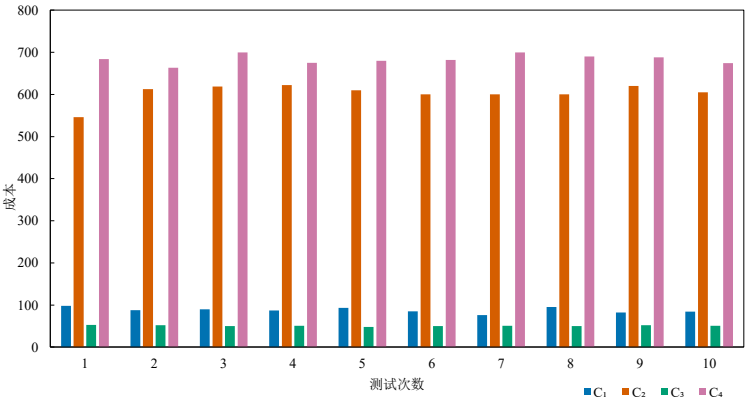


图 5 十次训练过程不同模型成本比较图

Fig.5 Comparison of the costs of different models in ten training processes

在全局成本测试图中， C_1 和 C_3 分别代表在 DQN 与带时间窗的 Dijkstra 策略下车辆从当前位置到充电站过程的成本； C_2 和 C_4 分别代表在 DQN 与带时间窗的 Dijkstra 策略下车辆在充电站产生的成本。每一次测试，分别对交通网络图中各个点进行一次导航测试，再将 36 个普通交通网络节点出发的电动汽车成本进行求和。

在多起点测试中，本文以若干普通交通网络节点作为出发点，对两种方法的成本进行逐点统计并绘制堆积柱状图，如图 6 所示。其中， D_1 表示电动汽车调用 DQN 模型进行实时充电导航时，在充电站所花费的总成本， D_2 表示在路程中所花费的总成本； J_1 表示电动汽车直接使用带时间窗的 Dijkstra 算法通过寻找最短路进行充电导航时，在充电站所花费的总成本， J_2 表示在路程中所花费的总成本。

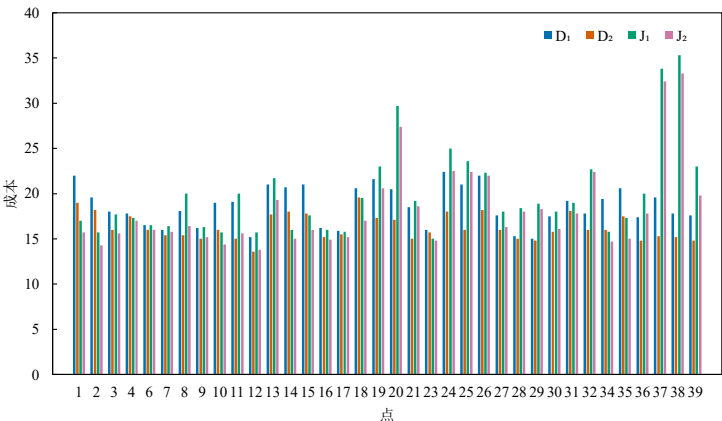


图 6 DQN 与带时间窗的 Dijkstra 算法下成本对比图

Fig.6 Cost Comparison Chart of DQN and Dijkstra's algorithm

对比分析每个点的成本，在部分起点上，有 14 个节点使用 DQN 模型的成本比带时间窗的 Dijkstra 模

型的成本略高,但其最大劣化幅度相对有限;相反,有 22 个节点表明,使用 DQN 模型比带时间窗的 Dijkstra 模型的成本更低,存在更显著的成本放大情形。该结果说明,DQN 的优势更集中体现在对高成本极端情形的规避能力上,能够有效降低由拥堵等待或不顺路绕行引发的长尾损失,从而使总体表现更稳健。

为了评估 DQN 策略在不同充电模式与时空场景下的泛化能力,在高峰与平峰时段分别对比四种决策方法。四种方法的计算规则统一如下: DQN 依据训练后的 Q 值在 Top-M 候选集中选择站点;带时间窗 Dijkstra 以 $O \rightarrow CS \rightarrow D$ 总行驶时间最小且时间窗可行为准;贪心最近站仅选择到站行驶时间 T_d 最小的可行站点;贪心最便宜站仅选择分时电价 π_j 最低的可行站点。四种方法均在同一套广义总成本公式、M/M/c 等待时间和充电时间参数下核算最终成本。

实验过程中,DQN 智能体选择的充电站(站 C)行驶路径流畅;带时间窗的 Dijkstra 与贪心最近选择的站 A 位于市中心高电价区(1.58 元/kWh),虽然行驶距离短,但需承受高峰排队($T_w=0.25h$);贪心最便宜选择的站 B 位于郊区(1.08 元/kWh),行驶距离显著增加($T_d=0.85h$)。

图 7 对比了四种方法在高峰与平峰时段的广义总成本。高峰时段,DQN 的广义总成本为 93.5 元,较带时间窗的 Dijkstra/贪心最近(99.4 元)降低 5.9%,较贪心最便宜(98.9 元)降低 5.5%;平峰时段,DQN 仍为最优(90.0 元),带时间窗的 Dijkstra/贪心最近(90.9 元),贪心最便宜为 97.4 元。

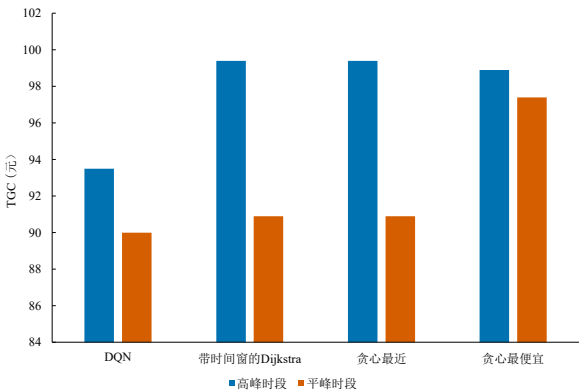


图 7 不同时段下各方法 TGC 对比

Fig.7 Comparison of TGC Across Different Methods Under Various Time Periods

2.3 算法运行效率分析

在本文的仿真环境中,全量充电站点总数量为 17 个,系统预筛保留候选站点数 10 个,搜索空间的压缩率约为 41.2%。即,在每一次推荐请求的处理过程中,系统减少了 41.2%的计算工程,极大改善了系统的响应处理能力。为了验证这种计算量的削减是否以牺牲决策质量为代价,本文基于包含 200 个独立任务的测试集,对预筛机制的一致性进行最终审计,具体性能审计汇总如表 6 所示。

表 6 预筛机制工程性能审计汇总表

Table 6 Summary Table of Engineering Performance Audit for the Pre-screening Mechanism					
审计指标	路网总站点数	预筛候选集规模	计算量压缩率	测试集样本量	一致性比例
参数数值	17	10	41.2%	200	100%

3 结论

1) 本文构建的广义总成本模型对用户时间价值系数表现出高度的敏感性与一致性, 证明模型能够准确模拟驾驶员在不同偏好下的决策逻辑, 有效将随机排队风险纳入了成本博弈体系。针对大规模城市路网导致强化学习动作空间过大、收敛困难的问题, 本文设计的双层架构经实证检验具备极高的工程落地价值。实验表明, 当预筛参数 $M=10$ 时, 算法将搜索空间压缩了约 41.2%, 在热启动状态下实现了 0.35ms 级的单次推理速度, 满足实时推荐系统的高并发需求。同时, 一致性检验显示该机制的 TOP-1 命中率高达 89.5%, 且未遗漏任何显著的全局最优解, 证明了利用行驶时间作为初筛代理指标高效合理。

2) 在 200 个独立随机任务的测试集上, 模型取得较高的 TOP-1 命中率 (89.5%) 与较低的平均相对后悔值 ($R_{rr}=0.0122$, 95%CI 为 [0.0060, 0.0183]), 且荷电状态安全约束违规率为 0%。典型算例复盘进一步说明: 智能体并非简单追求“最低电价”或“最短路径”, 而是能识别到站绕行与站内等待引入的时间价值损耗, 从而选择综合广义总成本更优的站点。

3) 本文仍存在一定局限。首先, M/M/c 排队模型采用短时稳态假设, 尚未显式刻画早晚高峰、节假日或突发事件下 λ_j 和 $c_j\mu$ 的快速波动, 后续可引入 M(t)/M/c、基于实时占用率的滚动更新或离散事件仿真增强动态鲁棒性。其次, 实验对象为南昌市中心城区 17 个充电站, 虽然可验证两阶段机制有效性, 但对包含数十至上百站点的大规模复杂路网仍需进一步测试, 并可加入 Dueling DQN、Rainbow 等深度强化学习基准进行横向比较。最后, 当前模型仍缺乏针对多车协同的联合优化机制, 未来可进一步考虑电网侧约束, 将电动汽车视为移动储能单元, 研究车网互动模式下的路径规划。

参考文献:

- [1] 行伟波, 张翊霆. 消费偏好差异与需求侧政策效率: 基于中国新能源汽车的分析 [J]. 软科学, 2026, 40 (5): 1-10.
Xing Weibo, Zhang Yiting. Differences in consumption preferences and the efficiency of demand-side policies: analysis based on Chinese NEV market [J]. Soft Science, 2026, 40 (5): 1-10.
- [2] Wang Yong, Zhou Jingxin, Sun Yaoyao, et al. Collaborative multidepot electric vehicle routing problem with time windows and shared charging stations [J]. Expert Systems with Applications, 2023, 219: 119654.
- [3] Çatay B, Sadati İ. An improved matheuristic for solving the electric vehicle routing problem with time windows and synchronized mobile charging/battery swapping [J]. Computers & Operations Research, 2023, 159: 106310.
- [4] Liu Weili, Gong Yuejiao, Chen Weineng, et al. Coordinated charging scheduling of electric vehicles: a mixed-variable differential evolution approach [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21 (12): 5094-5109.
- [5] Yao L, Lim W H, Tsai T S. A real-time charging scheme for demand response in electric vehicle parking station [J]. IEEE Transactions on Smart Grid, 2017, 8 (1): 52-62.
- [6] Lu Xinhui, Liu Zhaoxi, Ma Li, et al. A robust optimization approach for optimal load dispatch of community energy hub [J]. Applied Energy, 2020, 259: 114195.
- [7] Wang Bo, Zhao Dongbo, Dehghanian P, et al. Aggregated electric vehicle load modeling in large-scale electric power

- systems [J]. IEEE Transactions on Industry Applications, 2020, 56 (5): 5796-5810.
- [8] Tuhnitz F, Ebell N, Schlund J, et al. Development and evaluation of a smart charging strategy for an electric vehicle fleet based on reinforcement learning [J]. Applied Energy, 2021, 285: 116382.
- [9] Park K, Moon I. Multi-agent deep reinforcement learning approach for EV charging scheduling in a smart grid [J]. Applied Energy, 2022, 328: 120111.
- [10] Wang Guanhua, Yang Fang, Song Jian, et al. Optimization for dynamic laser inter-satellite link scheduling with routing: a multi-agent deep reinforcement learning approach [J]. IEEE Transactions on Communications, 2024, 72 (5): 2762-2778.
- [11] Zou Luyao, Munir M S, Hassan S S, et al. Imbalance cost-aware energy scheduling for prosumers towards UAM charging: a matching and multi-agent DRL approach [J]. IEEE Transactions on Vehicular Technology, 2024, 73 (3): 3404-3420.
- [12] Yu V F, Jodiawan P, Gunawan A. An Adaptive Large Neighborhood Search for the green mixed fleet vehicle routing problem with realistic energy consumption and partial recharges [J]. Applied Soft Computing, 2021, 105: 107251.
- [13] Comert S E, Yazgan H R. A new approach based on hybrid ant colony optimization-artificial bee colony algorithm for multi-objective electric vehicle routing problems [J]. Engineering Applications of Artificial Intelligence, 2023, 123: 106375.
- [14] Yang Senyan, Ning Lianju, Tong L C, et al. Optimizing electric vehicle routing problems with mixed backhauls and recharging strategies in multi-dimensional representation network [J]. Expert Systems with Applications, 2021, 176: 114804.
- [15] Liu Zhishuo, Zuo Xingquan, Zhou Mengchu, et al. Electric vehicle routing problem with variable vehicle speed and soft time windows for perishable product delivery [J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24 (6): 6178-6190.
- [16] 范泽敏, 吴翊恺, 王晨菡. 基于深度强化学习的智能车辆风险评估决策模型 [J]. 华东交通大学学报, 2026, 43 (1): 82-92.
- Fan Zemin, Wu Yikai, Wang Chenhan. Intelligent vehicle risk assessment decision model based on deep reinforcement learning [J]. Journal of East China Jiaotong University, 2026, 43 (1): 82-92.
- [17] 王强, 毕宇豪, 高超, 等. 融合多方信息的电动汽车充电负荷时空分布预测 [J]. 电力建设, 2025, 46 (6): 24-37.
- Wang Qiang, Bi Yuhao, Gao Chao, et al. Prediction of spatiotemporal distribution of electric vehicle charging load based on multi-source information [J]. Electric Power Construction, 2025, 46 (6): 24-37.
- [18] 梁士栋, 朱冰清. 考虑负荷分布均衡的电动汽车充电定价策略 [J]. 华东交通大学学报, 2022, 39 (6): 56-65.
- Liang Shidong, Zhu Bingqing. Charging pricing strategy for electric vehicles considering balanced time distribution of power load [J]. Journal of East China Jiaotong University, 2022, 39 (6): 56-65.
- [19] Wang Xiasen, MacKenzie D, Zhou Yang, et al. Predicted network equilibrium model of electric vehicles with stationary and dynamic charging infrastructure on the road network [J]. IEEE Intelligent Transportation Systems Magazine, 2022, 14 (2): 144-155.
- [20] Florio A M, Absi N, Feillet D. Routing electric vehicles on congested street networks [J]. Transportation Science,

2020, 55 (1): 238-256.

- [21] Davis N, Raina G, Jagannathan K. Grids versus graphs: partitioning space for improved taxi demand-supply forecasts [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22 (10): 6526-6535.
- [22] Wen Yanjie, Li Zhihong, Wang Xiaoyu, et al. Traffic demand prediction based on spatial-temporal guided multi graph Sandwich-Transformer [J]. Information Sciences, 2023, 643: 119269.
- [23] Wang Shengyou, Chen A, Wang Pinxi, et al. Predicting electric vehicle charging demand using a heterogeneous spatio-temporal graph convolutional network [J]. Transportation Research Part C: Emerging Technologies, 2023, 153: 104205.



第一作者: 郑逸凡 (2003—), 男, 硕士研究生, 研究方向为物流系统优化。



通信作者: 曾益 (1976—), 男, 副教授, 博士, 硕士生导师, 研究方向为运筹与系统优化、供应链网络可靠性。