基于空间感知的深度强化学习充电调度

吴素浓[1]，王　睿[2]

（1.华东交通大学交通运输工程学院，江西 南昌 330013；2.厦门大学管理学院，福建 厦门 361005）

摘要：针对电动汽车充电导航中行驶能耗、行驶时间、充电价格与等候时间难以统筹权衡的问题，提出一种基于空间感知注意力的深度 Q 网络充电调度方法（SCN-DQN）。在马尔可夫决策过程框架下，构建综合成本优化模型。首先，依据车辆荷电状态（SOC）筛选可达站点，并基于路网距离、充电价格和等候时间构建非支配候选集；其次，设计候选感知对称编码（CSE）与空间关联注意力（SRA）模块，提取站点相对特征并建模站点间关联，结合 Double DQN（DDQN）实现动作价值学习。最后，基于南昌市真实路网开展仿真实验，在四种单位时间成本情景下与五种对比策略进行比较。结果表明，SCN-DQN 在所有情景下均取得最低总成本，较 Dijkstra 最短路就近择站策略降低 0.92%～5.74%。

关键词：充电调度；深度强化学习；电动汽车；空间感知；注意力机制

中图分类号：TP 18　　**文献标志码：**A

Charging Scheduling for Electric Vehicles Based on Spatial-Cognition Station-Attention Deep Reinforcement Learning

Wu Sunong[1], Wang Rui[2]

(1. School of Transportation Engineering, East China Jiaotong University, Nanchang 330013, China；2. School of Management, Xiamen University, Xiamen 361005, China)

Abstract: To address the challenge of balancing driving energy consumption, travel time, charging prices, and waiting times in electric vehicle charging navigation, this study proposes a spatial-cognition network-based deep Q-network (SCN-DQN). A comprehensive cost optimization model is formulated within a Markov decision process framework. First, reachable charging stations are identified based on the vehicle's state of charge (SOC), and a non-dominated candidate set is constructed using road-network distance, charging price, and waiting time. Second, candidate-aware symmetric encoding (CSE) and spatial relational attention (SRA) modules are developed to extract relative station features and model inter-station relationships, with Double DQN (DDQN) employed for action-value learning. Finally, simulations are conducted on a real-world road network in Nanchang, China, comparing the proposed method with five strategies under four unit-time-cost scenarios. The results show that SCN-DQN achieves the lowest total cost across all scenarios, reducing costs by 0.92%–5.74% compared with the Dijkstra-based nearest-station strategy.

Key words: Charging scheduling; Deep reinforcement learning; Electric vehicle; Spatial cognition; Attention mechanism

收稿日期：2026−09−01

基金项目：江西省教育厅科学技术研究项目（GJJ2200630)

交通运输领域的能源消耗与碳排放是绿色低碳转型的重要议题。为落实碳达峰、碳中和目标，国家提出加快交通运输绿色化发展[1]。电动汽车凭借低排放、低能耗的优势，已成为交通领域节能减排的重要方向。截至 2025 年 12 月底，全国新能源汽车保有量达 4397 万辆，其中纯电动汽车 3022 万辆，占新能源汽车总量的 68.74%[2]。随着充换电基础设施持续完善，其建设与安全运行对电动汽车规模化应用具有重要意义[3]。

随着电动汽车保有量快速增长，充电基础设施的供需矛盾日益突出。充电站选址布局直接影响充电服务的可达性与便利性，已有研究以收益最大化为目标建立多等级选址模型[4]。在运营层面，不同站点的充电价格与等候时间存在差异，郑逸凡等[5]通过两阶段分层深度强化学习优化实时充电调度。此外，大规模电动汽车集中接入电网也会对电力系统安全稳定运行产生影响[6]。在实际充电过程中，车主容易忽略行驶电量、行驶时间、充电费用与等候时间之间的权衡，造成热门站点集聚、供需错配及充电总成本偏高[7]。因此，在多站点、多属性条件下为电动汽车选择综合成本较低的充电站，具有现实意义。

围绕上述问题，现有研究主要从静态优化与序贯决策两条路径展开。车辆路径问题为车辆调度和路径优化提供了经典框架[8]。早期研究多采用启发式或元启发式算法求解充电路径规划问题[9]，也有研究将深度强化学习与 Memetic 算法结合以提升求解效率[10]。此类方法通常依赖特定问题结构，在随机性和动态性较强的场景下适应性有限。深度强化学习则通过与环境交互学习决策策略，已应用于车辆路径规划[8]、机器人导航[11]等序贯决策问题，并为电动汽车充电调度提供了新的解决思路[7]。

深度强化学习以深度 Q 网络（DQN）为基础，通过深度神经网络逼近动作价值函数[12]，Double DQN 采用在线网络选择动作、目标网络评估价值，以缓解 Q 值过估计问题[13]。Dueling 架构则将状态价值与动作优势分离，进一步提升了价值估计的稳定性，这些方法为电动汽车充电调度提供了价值学习工具[14]。在充电调度领域，现有研究已在实时调度、需求响应、充放电协同与差异化服务等场景展开探索。面向实时调度，Wan 等[15]与李航等[16]分别构建了无模型的深度强化学习策略，Chen 等[17]进一步融合多元数据开展充电导航。Wen 等[18]从需求响应角度引入设备级强化学习，Xiao 等[19]考虑充放电双向能量流动，Zhang 等[20]则针对差异化充电诉求设计调度架构。随着研究深入，实时交通信息[21]与约束处理机制[22]也被显式纳入调度框架。总体来看，现有工作多关注时间维度上的充电负荷分配与功率调度，对多站点空间择站中的候选集构造、站点间相对比较与输入顺序一致性讨论仍不充分。

从表征角度看，多站点择站本质上是一个集合决策问题，站点输入顺序不应改变各物理站点的评价结果。Zaheer 等[23]提出的 Deep Sets 为集合数据的置换性质提供了表征框架，但其逐元素编码与池化结构对站点间交互的直接刻画较为有限。Lee 等[24]提出的 Set Transformer 利用注意力机制建模集合元素之间的关系，为站点间相对比较提供了可学习的实现途径，但注意力计算随站点数量呈二次增长，且未针对充电场景中的可达性约束与候选集筛选进行设计。在电动汽车充电推荐领域，Zhang 等[25]将注意力机制与多智能体强化学习结合，提出 RLCharge 算法，Su 等[26]则在交通网络、充电站与驾驶员联合建模中采用注意力多智能体方法。这些研究说明注意力机制能够刻画站点关联，但对单车择站中动态候选集、站点相对优劣及输入顺序一致性的针对性设计仍有进一步研究空间。

综上，现有研究尚有三方面不足。其一，站点表征多依赖固定顺序拼接，输入顺序的变化会直接改变网络输出，缺乏对置换等变性的显式保证。其二，候选集随车辆荷电状态与站点属性动态变化，但可达性筛选与非支配剪枝较少被纳入表征过程，网络只能在全部站点上被动学习约束，难以体现候选集的动态性。

其三，单车择站中的相对比较与多属性权衡尚未得到充分刻画，现有多智能体方法以协同收益为目标，更强调多车协同策略，而非单车站点间的相对优劣。针对上述不足，本文提出一种基于空间感知注意力的深度强化学习充电调度方法（Spatial-Cognition Network based Deep Q-Network，SCN-DQN），主要工作如下。

（1）将单车充电站选择问题建模为马尔可夫决策过程，构建综合考虑行驶能耗、行驶时间、充电及等候成本的优化模型，并依据车辆荷电状态与站点属性构建可达、非支配候选集，以缩减决策空间。

（2）设计候选感知对称编码（CSE）与空间关联注意力（SRA）组成的空间感知站点注意力网络。CSE 以候选集均值构造相对特征，并通过共享编码器生成具有置换等变性的站点表征。SRA 聚合候选站点间的信息，结合动作掩码估计各站点的动作价值。

（3）基于南昌市真实路网构建单车择站仿真场景，在四种单位时间成本下与五种对比策略进行比较，并通过成本分项、典型决策案例及消融实验评估方法性能与各模块的作用。结果表明，SCN-DQN 的总成本较 Dijkstra 最短路就近策略降低 0.92%~5.74%。

1 问题描述与模型构建

1.1 问题描述

在给定城市交通路网中，电动汽车从某一路网节点出发，需在满足最低到站荷电状态约束的前提下选择充电站并驶往该站。本文将充电导航建模为序贯择站过程：车辆在当前节点选择目标充电站，沿通往该站的最短路径行驶一个相邻路段，并在下一节点根据更新后的位置、电量及各站距离重新决策，直至到达目标站点。综合成本包括行驶能耗成本、行驶时间成本、充电成本和等候成本，其中前两项随实际行驶路段累积，后两项在到达充电站时产生。若车辆途中改变目标站点，已发生的行驶成本仍计入本次充电行程。因而，当前择站决策不仅影响下一路段的行驶成本，也影响后续可达站集合及最终充电成本。

将路网抽象为有向图$\mathcal{G}=(\mathcal{V},\mathcal{E})$，其中，$\mathcal{V}$ 和$\mathcal{E}$ 分别为节点集合与弧集合，弧权为道路长度，本文将每条无向道路按两个方向拆分为两条弧(i,j)与(j,i)，两者长度相同，即$l_{ij}=l_{ji}$。充电站集合记为$\mathcal{N}\subseteq\mathcal{V}$，其数量为$N=|\mathcal{N}|$，任一充电站表示为$n\in\mathcal{N}$，以$o\in\mathcal{V}$ 表示车辆的出发点，电动汽车初始荷电状态（state of charge，SOC）记为e_o。车辆当前位置至各充电站的路网最短距离通过最短路径算法计算，用于成本核算与站点可达性判断。

1.2 基本假设与符号约定

为便于建模，作如下假设：

（a）电动汽车行驶速度v在单次行程开始时按截断正态分布抽取，并在行程内保持不变；

（b）电动汽车每公里耗电量记为α，假设行驶耗电与行驶路径长度成正比；

（c）行驶途中平均电价为ρ_e，充电站n的充电单价记为λ_n；

（d）各充电站的充电价格λ_n与等候时间w_n在行程开始时确定，并在整个行程中保持不变。

本文主要符号及其含义见表 1。

表 1　主要符号说明

Tab.1　Main notations

符号	含义	单位	符号	含义	单位
$\mathcal{V}$	路网节点集合	—	v	行驶速度	km/h
M	路网节点数量	个	w_n	充电站 n 等候时间	h
$\mathcal{E}$	路网弧集合	—	w	各充电站等候时间向量	h
$\mathcal{N}$	充电站集合	—	d_t	时刻 t 至各充电站的最短距离向量	km
N	充电站数量	个	Δd_t	第 t 步的行驶距离	km
o	车辆出发节点	—	a_t	时刻 t 选择的充电站	—
e_o	初始荷电状态	—	Ω_t	时刻 t 的非支配候选站点集合	—
e_t	时刻 t 的荷电状态	—	r_t	时刻 t 的即时奖励	元
e_n^{end}	到达充电站 n 时的剩余荷电状态	—	δ_t	终止标记	—
$e_{\max}$	最高充电荷电状态	—	γ	折扣因子	—
$e_{\min}$	最低允许荷电状态	—	l_{ij}	弧 (i, j) 的长度	km
E_{total}	电池总容量	kWh	L_n	出发节点 o 至充电站 n 的最短路径长度	km
α	每公里耗电量	kWh/km	$d_{t,n}$	时刻 t 车辆当前位置至充电站 n 的最短距离	km
β	单位时间成本	元/h	C_n	选择充电站 n 的综合成本	元
ρ_e	行驶途中平均电价	元/kWh	s_t	时刻 t 的状态向量	—
λ_n	充电站 n 的充电价格	元/kWh	p_t	车辆当前位置的经纬度向量	—
λ	各充电站充电单价向量	元/kWh	x_{ij}	0-1 路径决策变量，1 代表沿弧 (i, j) 行驶，0 则不沿弧 (i, j) 行驶	—

1.3　充电导航模型

当用户选择充电站 n 时，以综合成本最小为目标，同时满足到站电量约束与路径连通性约束。

$$\min_{n\in\mathcal{N}} C_n = \alpha\rho_e L_n + \beta\frac{L_n}{v} + \lambda_n(e_{\max} - e_n^{\text{end}})E_{\text{total}} + \beta w_n \tag{1}$$

$$\text{s.t.}\quad\begin{cases} e_n^{\text{end}} \geq e_{\min} \\ \sum_{j:(o,j)\in\mathcal{E}} x_{oj} = 1 \\ \sum_{i:(i,n)\in\mathcal{E}} x_{in} = 1 \\ \sum_{i:(i,k)\in\mathcal{E}} x_{ik} = \sum_{j:(k,j)\in\mathcal{E}} x_{kj} \\ L_n = \sum_{(i,j)\in\mathcal{E}} x_{ij} l_{ij} \\ e_n^{\text{end}} = e_o - \dfrac{\alpha L_n}{E_{\text{total}}} \\ n \in \mathcal{N},\quad k \in \mathcal{V}\setminus\{o,n\},\quad x_{ij} \in \{0,1\} \end{cases} \tag{2}$$

式（1）表示优化目标函数，四项依次为行驶电量成本、行驶时间成本、充电成本与等候时间成本。式（2）第一项表示选择站点需要满足最低到站荷电状态约束，第二项表示车辆从出发节点选择一条道路驶出，第三项表示车辆通过一条道路进入目标充电站，第四项表示对于任一中间节点 k，进入该节点的路径流量与离开该节点的路径流量相等，第五项表示车辆从出发点到充电站 n 的总路径长度，第六项表示根据车辆初始荷电状态以及行驶过程中的电量消耗，到达充电站 n 时的剩余荷电状态。路径决策变量 x_{ij} 为二元变量，用于表示道路弧 (i, j) 是否被选取。

1.4 马尔可夫决策过程

考虑到车辆位置与荷电状态随行驶过程动态变化，将充电导航进一步建模为马尔可夫决策过程，以五元组 $(\mathcal{S},\mathcal{A},\mathcal{P},\mathcal{R},\gamma)$ 表示。

（1）状态空间 $\mathcal{S}$：状态由以下五个部分组成。

$$s_t = [e_t, p_t, d_t, \lambda, w] \tag{3}$$

式中，s_t 为时刻 t 的状态向量，e_t 为时刻 t 的荷电状态，p_t 为当前位置的经纬度向量，d_t 为车辆在时刻 t 到各充电站的最短距离向量，λ 和 w 分别为各充电站的充电价格和等候时间向量。

（2）动作空间 $\mathcal{A}$：总体动作空间由全部充电站构成，实际可选动作受到当前状态下筛选结果的约束。

$$a_t \in \Omega_t \subseteq \mathcal{A} \tag{4}$$

式中，a_t 为时刻 t 选择的充电站编号，Ω_t 为当前状态下经过可达性筛选和非支配筛选后形成的有效候选站点集合，其构造过程见 2.1 节。

（3）状态转移 $\mathcal{P}$：状态转移由路网连通性与行驶过程共同确定。设车辆在时刻 t 位于节点 i_t，选择目标站 a_t 后沿最短路径行驶一个相邻路段到达节点 i_{t+1}，转移过程为：

$$s_{t+1} = [e_{t+1}, p_{t+1}, d_{t+1}, \lambda, w] \tag{5}$$

$$e_{t+1} = e_t - \frac{\alpha \Delta d_t}{E_{\text{total}}} \tag{6}$$

$$d_{t+1,n} = \text{dist}(i_{t+1}, n), \quad n \in \mathcal{N} \tag{7}$$

$$\Delta d_t = l(i_t, i_{t+1}) \tag{8}$$

式（6）代表状态转移过程的电量消耗情况，式（7）中 $\text{dist}(i_{t+1}, n)$ 表示节点 i_{t+1} 与充电站点间 n 之间的路网最短距离，Δd_t 表示从节点 i_t 行驶到 i_{t+1} 的距离。

（4）奖励函数 $\mathcal{R}$：奖励取各成本分量的相反数。设时刻 t 选择的目标充电站为 a_t，则即时奖励为：

$$r_t = -[\alpha \rho_e \Delta d_t + \beta \frac{\Delta d_t}{v} + \delta_t (\lambda_{a_t} (e_{\max} - e_{t,a_t}) E_{\text{total}} + \beta w_{a_t})] \tag{9}$$

式中 r_t 为时刻 t 的即时奖励，Δd_t 为该步行驶距离，δ_t 为终止标记，车辆在本步到达目标充电站时其标记为 1，否则为 0，e_{t,a_t} 为车辆在时刻 t 选择充电站 a_t 后的预计到站荷电状态，w_{a_t} 为该站点的等候时间。

（5）折扣因子与回报：以折扣因子 $\gamma \in (0,1)$ 权衡当前与远期奖励，最大化累计期望折扣回报：

$$\max_{\pi} \mathbb{E}_{\pi}[\sum_{t \geq 0} \gamma^t r_t] \tag{10}$$

式中，π 为充电站选择策略，$\mathbb{E}_{\pi}$ 表示策略 π 下的期望运算，r_t 为时刻 t 的即时奖励。

2 基于 SCN-DQN 的空间感知注意力充电调度方法

在前述马尔可夫决策过程框架下，针对电动汽车多候选站点择站问题，本文提出一种基于空间感知注意力的深度 Q 网络充电调度方法（Spatial-Cognition Network-based Deep Q-Network，SCN-DQN）。该方法以车辆荷电状态、站点路网距离、充电价格和等候时间为决策依据，通过可达性筛选与非支配剪枝构建有效候选集，并结合站点相对特征编码与注意力交互，实现候选站点的动作价值估计，从而综合权衡行驶能耗、行驶时间、充电及等候成本。

SCN-DQN 由候选感知对称编码（Candidate-aware Symmetric Encoding，CSE）和空间关联注意力（Spatial Relational Attention，SRA）两个模块构成。CSE 以候选集均值为参照构造站点相对特征，并通过共享编码器生成逐站表征，SRA 则聚合候选站点间的关联信息，使动作价值估计兼顾站点自身属性及其与其他候选站点的相对关系。训练阶段采用经验回放与 Double DQN（DDQN）目标值更新机制优化网络参数。整体框架如图 1 所示。

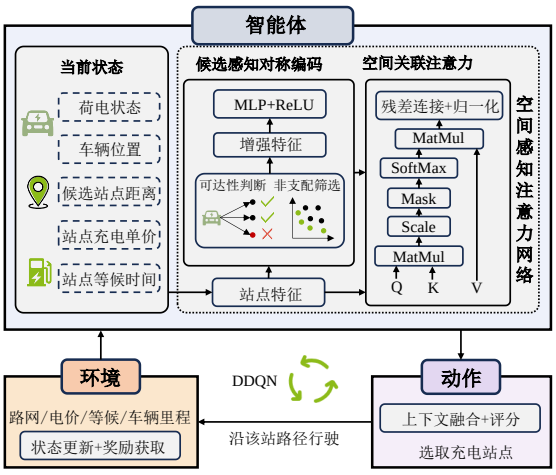


图 1 基于空间感知注意力的充电调度方法（SCN-DQN）总体框架

Fig.1 Overall framework of the spatial-cognition station-attention charging scheduling method (SCN-DQN)

2.1 候选感知对称编码（Candidate-aware Symmetric Encoding，CSE）

在多站点择站场景中，按照固定顺序拼接站点特征难以显式刻画站点间的相对竞争关系，且可能使网络输出受到输入排列的影响。为此，本文设计 CSE 模块：首先依据车辆荷电状态与路网距离筛除不可达站点，再根据路网距离、充电价格和等候时间构建非支配候选集，最后以候选集均值为参照提取站点相对特征，并通过共享编码器生成逐站表征。

设全部充电站集合为 $\mathcal{N}$，车辆在时刻 t 的荷电状态为 e_t，从当前位置至充电站 n 的路网最短距离为 $d_{t,n}$。根据最低到站荷电状态约束，可达站点集合定义为：

$$\mathcal{N}_t = \{n \in \mathcal{N} \mid e_t - \frac{\alpha d_{t,n}}{E_{\text{total}}} \geq e_{\min}\} \tag{11}$$

在可达站点集合内，若充电站 q 的路网距离、充电单价和等候时间均不高于充电站 n 的对应指标，且至少一项严格更低，则称充电站 q 支配充电站 n，记为 $q \succ n$。据此保留未被其他可达站点支配的候选集：

$$\Omega_t = \{n \in \mathcal{N}_t \mid \nexists q \in \mathcal{N}_t : q \succ n\} \tag{12}$$

设充电站 n 归一化后的特征向量为 $u_{t,n} = (d_{t,n}, \lambda_n, w_n)$，三个分量分别表示归一化后的路网距离、充电单价和预测等候时间。对于每个站点，将其原始特征与其相对于候选集均值的差值拼接，形成增强特征 $u_{t,n}$：

$$\bar{u}_t = \frac{1}{|\Omega_t|} \sum_{n \in \Omega_t} u_{t,n} \tag{13}$$

$$u_{t,n} = [u_{t,n}; u_{t,n} - \bar{u}_t] \tag{14}$$

式中，$\bar{u}_t$ 为时刻 t 的候选集均值向量，$|\Omega_t|$ 表示候选站点的数量。增强特征既保留站点自身的属性信息，又获得了该站点相对于当前候选集整体的差异。

随后，采用多层感知机对各站点的增强特征进行编码，得到统一维度的隐藏表征：

$$h_{t,n} = \sigma(W_2\sigma(W_1 u_{t,n} + b_1) + b_2) \tag{15}$$

式中，$h_{t,n}$ 为站点 n 编码后的表征，$\sigma(\cdot)$ 代表 ReLU 激活函数，W_1、W_2 为权重矩阵，b_1、b_2 为偏置向量。由于编码器对所有站点共享同一组参数且逐站独立作用，改变候选站点的输入顺序只会改变输出表征的排列，不会改变每个物理站点对应的表征内容，即该模块具有置换等变性。

2.2　空间关联注意力（Spatial Relational Attention，SRA）

CSE 主要提取各站点的独立表征，尚未显式建模候选站点间的相互作用。实际上，站点的相对吸引力不仅取决于自身属性，还与其他候选站点的距离、价格及等候时间有关。为此，本文将候选站点的隐藏表征组织为特征矩阵 H_t，采用多头自注意力机制聚合站点间的关联信息，并通过残差连接与层归一化增强训练稳定性，公式如下：

$$A_h = \text{softmax}(\frac{H_t W_Q^{(h)} (H_t W_K^{(h)})^{\mathrm{T}}}{\sqrt{d_h}}) H_t W_V^{(h)} \tag{16}$$

$$Z_t = \text{LN}(H_t + \text{Concat}(A_1, ..., A_{N_h}) W_O) \tag{17}$$

式中，A_h 为第 h 个注意力头的输出，$W_Q^{(h)}$、$W_K^{(h)}$ 和 $W_V^{(h)}$ 分别为该注意力头的查询、键和值投影矩阵，d_h 为查询和键向量的维度，上标 T 表示转置，softmax(·) 为归一化函数。Z_t 为经过空间关联注意力处理后的站点特征矩阵，N_h 为注意力头的数量，Concat(·) 表示特征拼接，W_O 为输出投影矩阵，LN(·) 为层归一化操作。

考虑到车辆荷电状态与当前位置同样影响择站决策，进一步采用上下文编码器将车辆状态映射为上下文向量 c_t，再与站点对应的增强表征 $z_{t,n}$ 融合，通过评分网络估计各候选站点的动作价值 $Q_\theta(s_t, n)$，最终智能体在有效候选集中选择动作价值最大的充电站 a_t^*，公式如下：

$$c_t = \sigma(W_3\sigma(W_4 s_t + b_3) + b_4) \tag{18}$$

$$Q_\theta(s_t, n) = W_{o'}\sigma(W_c[z_{t,n}; c_t] + b_c) + b_{o'} \tag{19}$$

$$a_t^* = \arg\max_{n \in \Omega_t} Q_\theta(s_t, n) \tag{20}$$

式（18）中，W_3、W_4 为上下文编码器的权重矩阵，b_3、b_4 为其偏置向量。式（19）中 $Q_\theta(s_t, n)$ 为在线网络对状态 s_t 下选择充电站 n 的动作价值估计，c_t 为车辆状态的上下文表征，$z_{t,n}$ 为站点 n 经过空间关联注意力处理后的特征，W_c 和 b_c 分别代表评分网络的权重参数与偏置向量，$W_{o'}$ 和 $b_{o'}$ 分别为输出权重向量和偏置。式（20）中 θ 表示在线网络的可训练参数，输出层采用线性映射以兼容负的动作价值。

2.3　DDQN 训练

标准 DQN 在动作选择与价值评估中使用同一网络，容易产生 Q 值过估计并引入策略偏差。为缓解这一问题，本文采用 Double DQN（DDQN），由在线网络选择下一状态的动作，并由目标网络评估该动作的

价值，其目标值为：

$$y_t = r_t + \gamma(1-\delta_t)Q_{\theta^-}(s_{t+1}, a_{t+1}^*) \tag{21}$$

其中 δ_t 为终止标记，若执行动作后进入终止状态，则 δ_t 为 1，否则为 0，采用均方误差构建网络回归损失：

$$\mathcal{L}(\theta) = \mathbb{E}[(y_t - Q_\theta(s_t, a_t))^2] \tag{22}$$

其中 θ 、θ^- 分别为在线网络与目标网络参数，目标网络每 C 步将在线参数同步一次。训练采用经验回放机制与 $\varepsilon-$贪心探索：从回放缓冲池中随机采样小批量样本进行更新，并随训练轮次将探索率由初值衰减至最小值。算法完整流程见算法 1。

算法 1 SCN-DQN 训练流程

Alg.1 Training process of SCN-DQN

输入：环境 $\mathcal{M}$；空间感知注意力网络 Q_θ；目标网络 Q_{θ^-}；回放缓冲池 $\mathcal{D}$；迭代轮数 T；批量大小 B

1. 初始化在线网络参数 θ 与目标网络参数 $\theta^- \leftarrow \theta$，清空回放缓冲池 $\mathcal{D}$
2. **for** $episode = 1, 2, ...T$ **do**
3. 初始化电动汽车状态 s_1
4. **for** each t **do**
5. 构造有效候选集 Ω_t，以概率 ε 随机选择动作 a_t；否则 $a_t = \arg\max_{n\in\Omega_t} Q_\theta(s_t, n)$
6. 执行动作 a_t，观测即时奖励 r_t 与下一状态 s_{t+1} 与终止标记 δ_t
7. 将样本 $(s_t, a_t, r_t, s_{t+1}, \delta_t)$ 存入 $\mathcal{D}$
8. 从 $\mathcal{D}$ 随机采样 B 条样本
9. 计算目标值 y_t：

$$y_t = r_t + \gamma(1-\delta_t)Q_{\theta^-}(s_{t+1}, a_{t+1}^*)$$

10. 计算回归损失 $\mathcal{L}(\theta)$，对 Q_θ 作梯度下降更新

$$\mathcal{L}(\theta) = \mathbb{E}[(y_t - Q_\theta(s_t, a_t))^2]$$

11. 每 C 步同步目标参数 $\theta^- \leftarrow \theta$；按训练轮次线性衰减探索率 ε
12. **end for**
13. **end for**

输出：训练完成的空间感知注意力网络 Q_θ

3 算例仿真与分析

3.1 仿真环境与路网构建

本文基于南昌市真实路网数据构建仿真环境。路网数据来源于 OpenStreetMap，采用网格采样方法提取 500 个路网节点，并根据空间分布随机选取 15 个节点作为候选充电站。图 2 展示了采样节点与候选充电站的空间分布，其中五角星表示充电站，蓝色圆点表示路网节点。

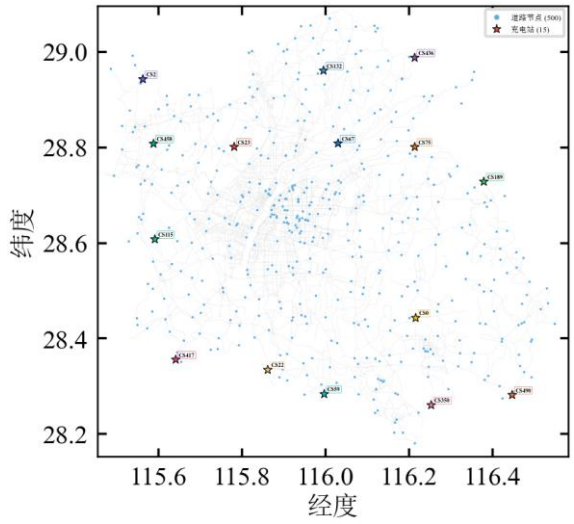


图 2 南昌市路网采样节点与充电站分布

Fig.2 Distribution of sampled road nodes and charging stations in Nanchang

表 2 列出了电动汽车及充电站的主要仿真参数。15 个候选充电站的充电价格分布的均值设为 2.0～3.2 元/kWh，且等候时间随充电价格升高而缩短，以模拟不同站点在价格与等候时间方面的差异，为多属性择站决策提供权衡条件。

表 2　仿真参数设置

Tab.2　Simulation parameter settings

参数	符号	取值/分布	单位	参数	符号	取值/分布	单位
电池总容量	E_{total}	40	kWh	行驶速度	v	$N(40,8^2)$	km/h
最高充电荷电状态	$e_{\max}$	0.80	–	行驶途中平均电价	ρ_e	U(0.60,1.00)	元/kWh
最低允许荷电状态	$e_{\min}$	0.15	–	单位时间成本	β	0/5/10/15	元/h
初始荷电状态	e_o	U(0.40,0.65)	–	充电站数量	N	15	个
每公里耗电量	α	0.18	kWh/km	路网节点数	M	500	个

3.2　空间感知注意力网络结构与训练参数

空间感知注意力网络的状态维度为 48，包含电量、经度、纬度以及 15 个充电站的距离、价格、等候时间。上下文编码器与站点编码器的隐藏层维度均为 64，注意力头数为 4，评分器将上下文特征与站点特征拼接后输出各充电站的 Q 值。SCN-DQN 训练的主要超参数如表 3 所示。

表 3　SCN-DQN 训练超参数

Tab.3　SCN-DQN training hyper-parameters

参数	取值	参数	取值
隐藏层维度	64	折扣因子	0.95
注意力头数	4	经验回放容量	50000
学习率	5×10^{-4}	批量大小	128
初始 ε	0.95	最小 ε	0.01
目标网络更新间隔	20	训练轮数（Epoch）	500

4　实验结果与分析

4.1　对比算法

为评估 SCN-DQN 在充电站选择任务中的性能，选取五种对比策略。各策略采用相同的路网、车辆初始状态及站点充电价格、等候时间等基础设置，具体如下：

（1）Dijkstra 算法：计算车辆当前位置至各充电站的路网最短距离，并以行驶距离最短为选站依据，作为就近择站基准。

（2）价格贪心策略（Greedy-Price）：选择充电单价最低的站点，用于评估仅考虑充电价格的决策效果。

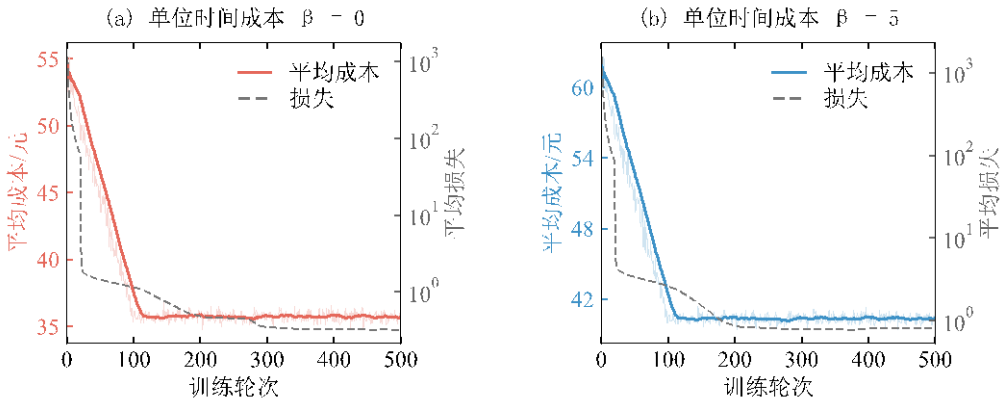
（3）等候时间贪心策略（Greedy-Wait）：选择等候时间最短的站点，用于评估仅考虑等候时间的决策效果。

（4）加权求和策略（Weighted-Sum）。对路网距离、充电价格和等候时间进行归一化处理，并依据预设权重计算站点评分，选择可达站点中评分最低者，以评估基于固定权重的多指标融合策略。其中路网距离、充电价格与等候时间的预设权重分别为 1、1 与 β/5。

（5）随机策略（Random）：以等概率随机选择目标站点，作为不依据站点属性进行择优决策的参照。

4.2　算法训练与收敛性分析

图 3 给出不同单位时间成本下 SCN-DQN 的训练曲线。各情景下的平均成本随训练轮数增加总体呈下降趋势，并在约 100 轮后逐渐趋于平稳。单位时间成本为 0 元/h 时的收敛成本最低，为 15 元/h 时最高，反映出时间成本系数对综合成本的影响。与此同时，训练损失总体下降并逐渐稳定，表明网络训练过程趋于平稳。



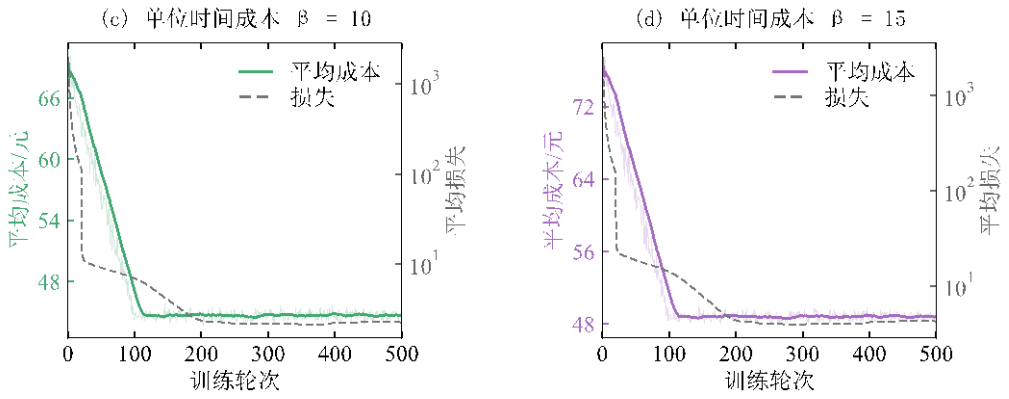


图 3 不同单位时间成本下的训练曲线

Fig.3 Training curves under different unit time cost

4.3 不同单位时间成本下的性能对比

为评估不同时间成本系数下的充电调度性能，将单位时间成本分别设为 0、5、10 和 15 元/h。各情景均采用相同的 485 个非充电站路网节点作为出发点，并保持车辆初始状态及站点参数的生成规则一致。各算法的总成本如表 4 所示。

表 4 不同单位时间成本下各算法的总成本（元）

Tab.4 Total costs of different algorithms under different unit time cost (CNY)

算法	β =0	β =5	β =10	β =15
Dijkstra	17946.27	19797.72	21649.18	23500.64
Greedy-Price	21625.51	25807.47	29989.43	34171.39
Greedy-Wait	34114.17	37711.19	41308.20	44905.21
Weighted-Sum	18469.21	22484.78	31764.11	38581.98
Random	27952.74	31845.77	35738.80	39631.84
SCN-DQN	**16915.32**	**19161.47**	**21286.67**	**23285.31**

由表 4 可知，SCN-DQN 在四种情景下的总成本均低于五种对比策略。单位时间成本为 5 元/h 时，其总成本为 19161.47 元，较 Dijkstra 算法降低 3.21%。随着单位时间成本由 0 元/h 增至 15 元/h，SCN-DQN 相对 Dijkstra 算法的成本降幅由 5.74%逐渐收窄至 0.92%。这主要是因为行驶距离与行驶时间具有较强关联，时间成本权重提高时，Dijkstra 的就近择站决策更接近低时间成本选择，而在较低时间成本情景下，综合考虑站点价格所带来的收益更为明显。

为进一步分析成本差异的来源，表 5 给出了单位时间成本为 5 元/h 时 485 个评估场景的总成本及各项指标的单场景平均值。与 Dijkstra 算法相比，SCN-DQN 的平均路径能耗增加 0.42kWh，平均等候时间增加 4.46min，但平均所选站点充电单价由 2.50 元/kWh 降至 2.29 元/kWh，单场景平均充电成本降低 2.35 元，最终总成本减少 636.25 元。价格贪心策略虽然选择了单价更低的站点，但较高的路径能耗使其平均充电成本仍高于 Dijkstra 算法。等候时间贪心策略虽显著缩短了等候时间，但较高的路径能耗与充电成本使其总成本最高。上述结果表明，SCN-DQN 的成本优势主要来自对行驶、充电与等候成本的综合权衡，而非单一指标的局部优化。

表 5　β=5 时各算法的总成本与分项指标

Tab.5　Total cost and component metrics of different algorithms at β=5

算法	总成本(元)	平均等待时间(分钟)	平均路径能耗(kWh)	所选充电站平均电价 (元/kWh)	平均充电成本(元)
Dijkstra	19797.72	18.24	**3.13**	2.50	34.47
Greedy-Price	25807.47	29.24	8.54	**1.96**	37.72
Greedy-Wait	37711.19	**5.09**	9.71	3.07	62.54
Weighted-Sum	22484.78	16.60	5.08	2.37	37.18
Random	31845.77	17.43	9.08	2.55	50.33
SCN-DQN	**19161.47**	22.70	3.55	2.29	**32.12**

为进一步说明单位时间成本变化对择站决策的影响，以出发节点 145 为例。表 6 列出了该案例中相关充电站的成本构成，表 7 给出了各策略在四种单位时间成本情景下的选站结果及对应总成本。

表 6　出发节点 145 所涉及充电站的成本构成

Tab.6　Cost profiles of charging stations selected from origin node 145

充电站	行驶能耗成本（元）	驾驶时间(h)	充电成本(元)	等待时间(h)
1	4.03	0.53	26.60	0.52
8	10.49	1.39	54.67	0.26
12	10.05	1.33	62.24	0.09
15	3.15	0.42	31.18	0.23

表 7　出发节点 145 在不同单位时间成本下的选站结果与总成本（元）

Tab.7　Station choices and total costs under different β values

算法	所选充电站	β=0	β=5	β=10	β=15
Dijkstra	15	34.33	37.59	40.84	44.10
Greedy-Price	1	30.63	35.89	41.16	46.42
Greedy-Wait	12	72.28	79.38	86.48	93.58
Weighted-Sum	1->15->15->12	30.63	37.59	40.84	93.58
Random	8	65.16	73.37	81.59	89.80
SCN-DQN	1->1->15->15	30.63	35.89	40.84	44.10

由表 6 可知，站点 1 的充电成本低于站点 15，但其行驶电量成本、行驶时间和等候时间均较高。根据两站的原始成本数据计算，当单位时间成本约为 9.22 元/h 时，两站的综合成本相等。由表 7 可知，分别在单位时间成本为 5 元/h 和 10 元/h 的情景下训练的 SCN-DQN 依次选择站点 1 和站点 15，与两站综合成本排序的变化一致。尽管站点 12 的等候时间最短，但其行驶电量成本和充电成本较高，因而在该案例的各情景下综合成本均较高。该案例表明，充电站的选择需要综合权衡行驶、充电与等候等成本因素。

4.4　消融实验

为分析候选感知对称编码（CSE）与空间关联注意力（SRA）对决策的作用，在单位时间成本为 5 元

/h 的情景下设置三种结构变体与完整 SCN-DQN 进行比较。各模型均训练 500 轮且测试场景保持一致。

（1）SCN-DQN（wo CSE）：移除候选集均值参照、相对特征及双层共享站点编码器，改用原始站点特征的单层线性映射，同时保留 SRA，以评估候选相对编码的作用。

（2）SCN-DQN（wo SRA）：保留 CSE，移除多头自注意力机制，使各站点编码后不再聚合其他候选站点的信息，以评估站点间关联建模的作用。

（3）SCN-DQN（Plain MLP）：采用普通多层感知机直接处理按固定顺序拼接的状态特征，同时移除 CSE 与 SRA，评价时仍使用相同的候选动作掩码，以比较整体网络结构的效果。

表 8　消融实验结果（β=5）

Tab.8　Ablation results (β=5)

算法	总成本	相对 Dijkstra 成本变化率（%）
SCN-DQN	**19161.47**	**-3.21**
SCN-DQN（wo CSE）	19219.61	-2.92
SCN-DQN（wo SRA）	19200.54	-3.02
SCN-DQN（Plain MLP）	19528.80	-1.36

表 8 给出了消融实验结果。完整 SCN-DQN 的总成本最低，为 19161.47 元，较 Dijkstra 算法降低 3.21%。移除 CSE 后，总成本增至 19219.61 元，降幅降至 2.92%；移除 SRA 后，总成本增至 19200.54 元，降幅降至 3.02%。采用 Plain MLP 后，总成本进一步增至 19528.80 元，降幅仅为 1.36%。上述结果表明，在该测试情景下，CSE 与 SRA 均有助于降低综合成本，二者结合的完整模型取得了最佳表现。

5　结论

本文针对电动汽车充电站选择中的多属性权衡问题，提出了一种基于空间感知注意力的深度强化学习充电调度方法（SCN-DQN），主要结论如下：

（1）将单车充电站选择问题建模为马尔可夫决策过程，构建综合考虑行驶能耗、行驶时间、充电及等候成本的优化模型，为多因素约束下的充电站选择提供了建模基础。

（2）设计了由候选感知对称编码（CSE）与空间关联注意力（SRA）组成的空间感知站点注意力网络。CSE 通过候选集均值构造站点相对特征，并利用共享编码器生成具有置换等变性的站点表征，SRA 通过聚合候选站点间的信息，使动作价值估计兼顾自身属性及其相对关系，从而增强模型的相对比较能力。

（3）基于南昌市真实路网开展实验。结果表明，在四种单位时间成本情景下，SCN-DQN 的总成本均低于五种对比策略，较 Dijkstra 就近择站策略降低 0.92%～5.74%，验证了该方法具有较好的综合效果。

（4）消融实验表明，移除 CSE 或 SRA 均会导致总成本上升，完整模型在所测试情景下取得最低总成本，验证了两个模块对调度性能的贡献。

本文仍存在一定局限。首先，当前模型采用相对静态的路网信息，充电价格与等候时间作为外生参数给定，尚未刻画站点排队过程及车辆选择对等候时间的动态反馈，且未将实际充电时长纳入时间成本。可融合动态交通信息、实时站点状态及排队反馈，提升模型对动态环境的适应能力。其次，本文研究充电场景下的单车决策，未涉及多车并发调度及推理时延。可拓展至多车、多站协同决策，并在车载平台开展推理性能测试。此外，可将动态电价等不确定因素纳入模型，以增强方法在实际场景中的适用性与鲁棒性。

参考文献：

[1] 谭显东, 刘俊, 徐志成, 等. “双碳”目标下“十四五”电力供需形势 [J]. 中国电力, 2021, 54(5): 1-6.
Tan Xiandong, Liu Jun, Xu Zhicheng, et al. Power supply and demand situation during the 14th Five-Year Plan under the “dual carbon” goal. Electric Power, 2021, 54(5): 1-6.

[2] 中华人民共和国公安部. 全国机动车保有量达 4.69 亿辆 驾驶人达 5.59 亿名 [EB/OL]. https://www.mps.gov.cn/n2254314/n6409334/c10383533/content.html, 2026-01-27.
Ministry of Public Security of the People's Republic of China. National motor vehicle ownership reached 469 million, drivers reached 559 million [EB/OL]. https://www.mps.gov.cn/n2254314/n6409334/c10383533/content.html, 2026-01-27.

[3] 于东民, 杨超, 蒋林洳, 等. 电动汽车充电安全防护研究综述 [J]. 中国电机工程学报, 2022, 42(6): 2145-2164.
Yu Dongmin, Yang Chao, Jiang Linru, et al. Review of research on electric vehicle charging safety protection. Proceedings of the CSEE, 2022, 42(6): 2145-2164.

[4] 宋志成, 王勋, 伦利, 等. 基于收益最大化的电动汽车充电站选址规划 [J]. 华东交通大学学报, 2014, 31(1): 50-55.
Song Zhicheng, Wang Xun, Lun Li, et al. Location planning of electric vehicle charging stations based on profit maximization. *Journal of East China Jiaotong University*, 2014, 31(1): 50-55.

[5] 郑逸凡, 曾益, 易心雨, 等. 基于两阶段分层 DQN 的电动汽车实时充电调度策略[J]. 华东交通大学学报, 2026, 1-14.
Zheng Yifan, Zeng Yi, Yi Xinyu, et al. Real-time electric vehicle charging scheduling strategy based on two-stage hierarchical DQN. Journal of East China Jiaotong University, 2026, 1-14.

[6] 胡泽春, 宋永华, 徐智威, 等. 电动汽车接入电网的影响与利用 [J]. 中国电机工程学报, 2012, 32(4): 1-10+25.
Hu Zechun, Song Yonghua, Xu Zhiwei, et al. Impacts and utilization of electric vehicles integration into power grids. Proceedings of the CSEE, 2012, 32(4): 1-10+25.

[7] 张延宇, 饶新朋, 周书奎, 等. 基于深度强化学习的电动汽车充电调度算法研究进展 [J]. 电力系统保护与控制, 2022, 50(16): 179-187.
Zhang Yanyu, Rao Xinpeng, Zhou Shukui, et al. Research progress on electric vehicle charging scheduling algorithms based on deep reinforcement learning. *Power System Protection and Control*, 2022, 50(16): 179-187.

[8] 杨笑笑, 柯琳, 陈智斌. 深度强化学习求解车辆路径问题的研究综述 [J]. 计算机工程与应用, 2023, 59(5): 1-13.
Yang Xiaoxiao, Ke Lin, Chen Zhibin. Review of deep reinforcement learning for solving vehicle routing problems. Computer Engineering and Applications, 2023, 59(5): 1-13.

[9] 周鲜成, 刘长石, 周开军, 等. 时间依赖型绿色车辆路径模型及改进蚁群算法 [J]. 管理科学学报, 2019, 22(5): 57-68.
Zhou Xiancheng, Liu Changshi, Zhou Kaijun, et al. Time-dependent green vehicle routing model and improved ant colony algorithm. Journal of Management Sciences in China, 2019, 22(5): 57-68.

[10] 周雅兰, 廖易天, 粟筱, 等. 深度强化学习 Memetic 算法求解取送货车辆路径问题 [J]. 计算机科学与探索, 2024, 18(3): 818-830.
Zhou Yalan, Liao Yitian, Su Xiao, et al. Deep reinforcement learning Memetic algorithm for solving pickup and delivery vehicle routing problem. Journal of Computer Science and Exploration, 2024, 18(3): 818-830.

[11] 熊李艳, 舒垚淞, 曾辉, 等. 基于深度强化学习的机器人导航算法研究 [J]. 华东交通大学学报, 2023, 40(1): 67-74.
Xiong Liyan, Shu Yaosong, Zeng Hui, et al. Research on robot navigation algorithm based on deep reinforcement learning. Journal of East China Jiaotong University, 2023, 40(1): 67-74.

[12] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[13] Van H, Guez A, Silver D. Deep reinforcement learning with double q-learning[C]//AAAI. 2016, 30(1): 2094-2100.

[14] Wang Z, Schaul T, Hessel M, et al. Dueling network architectures for deep reinforcement learning[C]//ICML. 2016: 1995-2003.

[15] Wan Z, Li H, He H, et al. Model-free real-time EV charging scheduling based on deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2019, 10(5): 5246-5257.

[16] 李航, 李国杰, 汪可友. 基于深度强化学习的电动汽车实时调度策略 [J]. 电力系统自动化, 2020, 44(22): 161-167.
Li Hang, Li Guojie, Wang Keyou. Real-time scheduling strategy for electric vehicles based on deep reinforcement learning. Automation of Electric Power Systems, 2020, 44(22): 161-167.

[17] Chen L, Yang F, Wu S, et al. Electric vehicle charging navigation strategy based on data driven and deep reinforcement learning[C]// CCEAI. 2021: 16-23.

[18] Wen Z, O'Neill D, Maei H. Optimal demand response using device-based reinforcement learning[J]. IEEE Transactions on Smart Grid, 2015, 6(5): 2312-2324.

[19] Xiao Q, Zhang R, Wang Y, et al. A deep reinforcement learning based charging and discharging scheduling strategy for electric vehicles[J]. Energy Reports, 2024, 12:4854-4863.

[20] Zhang A, Liu Q, Liu J, et al. CASA: cost-effective EV charging scheduling based on deep reinforcement learning[J]. Neural Computing & Applications, 2024, 36(15).
[21] 邢强, 陈中, 冷钊莹, 等. 基于实时交通信息的电动汽车路径规划和充电导航策略 [J]. 中国电机工程学报, 2020, 40(2): 534-550.
Xing Qiang, Chen Zhong, Leng Zhaoying, et al. Electric vehicle route planning and charging navigation strategy based on real-time traffic information. Proceedings of the CSEE, 2020, 40(2): 534-550.
[22] Chen G, Yang L, Cao X. A deep reinforcement learning-based charging scheduling approach with augmented Lagrangian for electric vehicles[J]. Applied Energy, 2025, 378.
[23] Zaheer M, Kottur S, Ravanbakhsh S, et al. Deep sets[J]. Advances in neural information processing systems, 2017, 30.
[24] Lee J, Lee Y, Kim J, et al. Set transformer: A framework for attention-based permutation-invariant neural networks[C]//International conference on machine learning. PMLR, 2019: 3744-3753.
[25] Zhang W, Liu H, Xiong H, et al. RLCharge: Imitative multi-agent spatiotemporal reinforcement learning for electric vehicle charging station recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(6): 6290-6304.
[26] Su S, Li Y, Yamashita K, et al. Electric vehicle charging guidance strategy considering “traffic network-charging station-driver” modeling: a multiagent deep reinforcement learning-based approach[J]. IEEE Transactions on Transportation Electrification, 2023, 10(3): 4653-4666.



第一作者：吴素浓（1979—），女，副教授，研究方向包括路径优化、文本挖掘、大数据挖掘等。



通信作者：王睿（2002—），男，博士研究生，研究方向包括运筹优化、大数据挖掘等。