基于分层深度强化学习的电动汽车实时充电引导策略

陆文韬[1,2]，窦胜[1,2]，陈良亮[1,2]，杨凤坤[1,2]，周瑞超[3]
（1.国网电力科学研究院有限公司，江苏 南京 211100；
2.国电南瑞南京控制系统有限公司，江苏 南京 211100；
3.天津大学 智能电网教育部重点实验室，天津 300072）

摘要：为了实现电动汽车的实时充电引导以及提高充电站的充电效率，提出了一种基于分层深度强化学习的电动汽车实时充电引导策略。考虑车-站-路多元主体的相互耦合特性，基于电动汽车与充电站、配电网和交通路网的特征信息构建双层电动汽车充电导航模型。将上述模型解耦成双层有限马尔可夫决策过程网络架构，上层网络评估和推荐充电站，并将最优选择结果传递给下层网络，下层网络为用户规划行驶路径。采用基于彩虹框架的深度Q网络算法求解上述双层决策过程。最后在某特定城市区域进行仿真验证，结果表明，与无序引导方法相比，所提方法可以减少用户时间成本和节省用户费用，且能够保证配电网安全运行。

关键词：电动汽车；实时充电引导；推荐充电站；规划行驶路径；双层深度强化学习；深度Q网络算法
中图分类号：TM73 **文献标识码**：A **DOI**：10.19457/j.1001-2095.dqcd26280

Real-time Charging Guidance Strategy for Electric Vehicles Based on Hierarchical Deep Reinforcement Learning

LU Wentao[1,2]，DOU Sheng[1,2]，CHEN Liangliang[1,2]，YANG Fengkun[1,2]，ZHOU Ruichao[3]
（1.*State Grid Electric Power Research Institute*，*Nanjing* 211100，*Jiangsu*，*China*；
2.*NARI Technology Nanjing Control Systems Co.*，*Ltd.*，*Nanjing* 211100，*Jiangsu*，*China*；
3.*Key Laboratory of Smart Grid of Ministry of Education*，*Tianjin University*，*Tianjin* 300072，*China*）

Abstract：To realize the real-time charging guidance of electric vehicles and improve the charging efficiency of charging stations，a real-time charging guidance strategy for electric vehicles based on hierarchical deep reinforcement learning was proposed. Considering the mutual coupling characteristics of vehicle-station-road multiple agents，a double-layer electric vehicle charging navigation model was constructed based on the characteristic information of electric vehicles，charging stations，distribution networks and transportation networks. The above-mentioned model was decoupled into a two-layer finite Markov decision process network architecture，the upper network evaluated and recommended charging stations，and the optimal selection result were passed to the lower network. The lower network planed the driving path for the user. The deep Q-network algorithm based on rainbow framework was used to solve the above-mentioned two-layer decision-making process. Finally，the simulation results in a specific urban area show that compared with the disorderly guidance method，the proposed method can reduce the user time cost and save the user cost，and ensure the safe operation of the distribution network.

Key words：electric vehicle（EV）；real-time charging guidance；recommending charging station；planning driving path；two-layer deep reinforcement learning（DRL）；deep Q-network algorithm

电动汽车（electric vehicle，EV）作为“碳达峰、碳中和”能源转型路径的重要组成部分，得到了各级政府的积极推广，截至2023年底，我国新能源汽车保有量达2 041万辆，车桩比为2.4∶1[1]。

基金项目：国家电网有限公司科技项目（5400-202312239A-1-1-ZN）
作者简介：陆文韬（1999—），男，硕士，主要研究方向为电动汽车与电网互动技术，Email：luwentao@sgepri.sgcc.com.cn
通讯作者：陈良亮（1975—），男，博士，正高级工程师，主要研究方向为电动汽车充换电技术，Email：chenliangliang@sgepri.sgcc.com.cn

然而,充电基础设施建设步伐远落后于电动汽车爆发速度,造成了用户充电效率低下、充电便利性不足等问题,严重影响了用户的驾乘体验以及车辆的进一步普及[2]。为此,合理高效的电动汽车充电引导是实现能源交通友好融合的必要前提。

目前为止,诸多研究已经深入开展了电动汽车智能充电引导,旨在为用户的日常行驶和能源补给提供便捷的决策辅助。针对电动汽车用户出行成本影响分析,文献[3-5]根据时间与费用消耗制定最优行驶与充电引导方案。文献[3]统筹考虑交通-电气耦合信息以及实时电价策略,提出基于动态电价激励的EV充电引导方案,降低用户出行的时间与充电成本。进一步,文献[4]通过为快充需求用户规划经济行驶路径与推荐最优充电站,实现用户成本降低的同时提高快充站的运营效率。文献[5]分析了交通流延迟性对电动汽车用户出行成本的影响,通过构建EV充电决策模型实现配电-交通耦合系统协同经济运行。

文献[6-7]从增强EV充电网络智能化管理水平,综合分析"车路网耦合"环境下的能耗、成本和通行效率,为车主提供高效的充电策略。文献[6]分析了"车-站-路-网"耦合条件下影响EV能耗的相关因素,将充电导航问题形式化为两阶段随机优化问题,为EV车主提供便捷的充电导航方案。文献[7]考虑了实时交通信息对用户决策的影响,建立了基于"车-站-网"耦合的充电引导模型,优化用户出行成本、充电站聚集充电负荷以及路网通行效率。

此外,文献[8-10]考虑充电运营商服务水平,通过改善充电便捷性与经济性,提升用户的充电体验。基于实时路网状态特征信息,文献[8-9]在确保满足用户充电需求和充电站利用率的条件下,提出多决策方案结合的充电导航策略。进一步,文献[10]提出电动汽车分布式引导框架,该框架协调EV用户、快速充电站、EV分配中心和配电网运营商等多方面主体,优化电动汽车车主费用以及快速充电站的充电效率。

上述研究为深入理解电动汽车充电导航的决策机制奠定了研究基础,然而传统的数学建模和启发式算法在面对大规模交通-电气网络时,往往面临计算效率低下和实时性不足等问题。

近年来,机器学习得到快速发展,尤其是深度强化学习(deep reinforcement learning,DRL),验证了其在处理EV充电引导问题上的可行性。为此,文献[11-14]聚焦如何提高充电用户引导过程中的在线应用效果。文献[11]通过确定最短充电路线模型来获取车网的关键特征,并将其构建为有限马尔可夫决策过程(finite Markov decision process,FMDP)来进行最优充电导航策略的实时学习。在文献[12]中,引入行为经济学的助推理论来优化EV车主的充电计划,通过孪生延迟深度确定性策略梯度算法对单辆EV进行实时充电引导。文献[13]基于多层网络理论的快速充电导航策略,利用耦合网络加权定价方法解决电动汽车快速充电导航问题。文献[14]提出换电站实时调度策略,基于蒙特卡罗策略梯度强化学习方法实现换电站实时调度,优化了换电站的充放电策略和响应电池数量。

进一步,为了优化充电引导的多主体间的协作效率,文献[15-17]基于深度强化学习算法设计了考虑充电需求竞争的在线充电引导方案。文献[15]提出了在线电动汽车充电引导算法,采用集中训练和去中心化执行的Actor-critic算法框架,实现了大规模电动汽车的高效充电导航。而文献[16]通过构建双时间尺度耦合框架,设计高效的电动汽车在途充电引导策略,优化主体间的协作效率,确保充电网络的高效运行。文献[17]基于深度强化学习方法,解决配电系统中电力与交通系统动态交互的贯序协同优化问题,显著提升多主体在充电引导过程中的协作效率。

尽管上述基于DRL方法为电动汽车充电引导提供了多种建模思路和先进的实时求解算法,然而面对用户行为多样性和车网环境多变性时,需要满足环境安全运行前提下兼顾用户经济性引导。

综上,考虑到这一领域电动汽车智能充电引导策略,仍然存在两个不足:

第一,DRL算法在优化目标与安全约束之间的平衡问题[18]。目前传统人工智能方法在训练和决策过程中往往忽视了安全约束条件带来的影响,导致决策结果与实际安全运行要求存在偏差,难以在复杂的电网环境中精准地平衡控制目标与安全限制。

第二,现有DRL算法在提供端到端解决方案方面的不足[19]。许多DRL算法专注于单一任务,例如路径规划或站点选择,而缺乏一个能够同时

处理路径规划和站点选择的综合性解决方案。在复杂的动态环境中,单一任务的优化往往无法满足系统整体性能的最优化需求。

综上,本文提出基于分层深度强化学习(hierarchical deep reinforcement learning,HDRL)的电动汽车充电引导策略,上层网络评估和选择最优充电站,下层网络优化给定充电站的最优充电行驶路径,并基于彩虹算法架构进行求解,最后通过实际城市网络拓扑进行仿真验证。

1　引导结构

基于HDRL的电动汽车充电引导整体框架如图1所示,本文构建的电动汽车充电引导策略是基于分层强化学习的双层决策架构。该架构主要由上、下两层构成,当电动汽车发出充电请求时,上层网络通过深度强化学习算法进行充电站的决策,以确定最佳的充电站点,上层智能体依据环境提供的状态信息和反馈奖励,优化动作选择策略。一旦确定了最优充电站,相关信息被传递至下层网络,由下层智能体负责制定充电行驶路径,下层采用与上层相同的深度强化学习算法。所提引导策略遵循顺序调度模型,智能体依次为每辆电动汽车生成充电与路径决策,直至所有车辆调度工作完成。

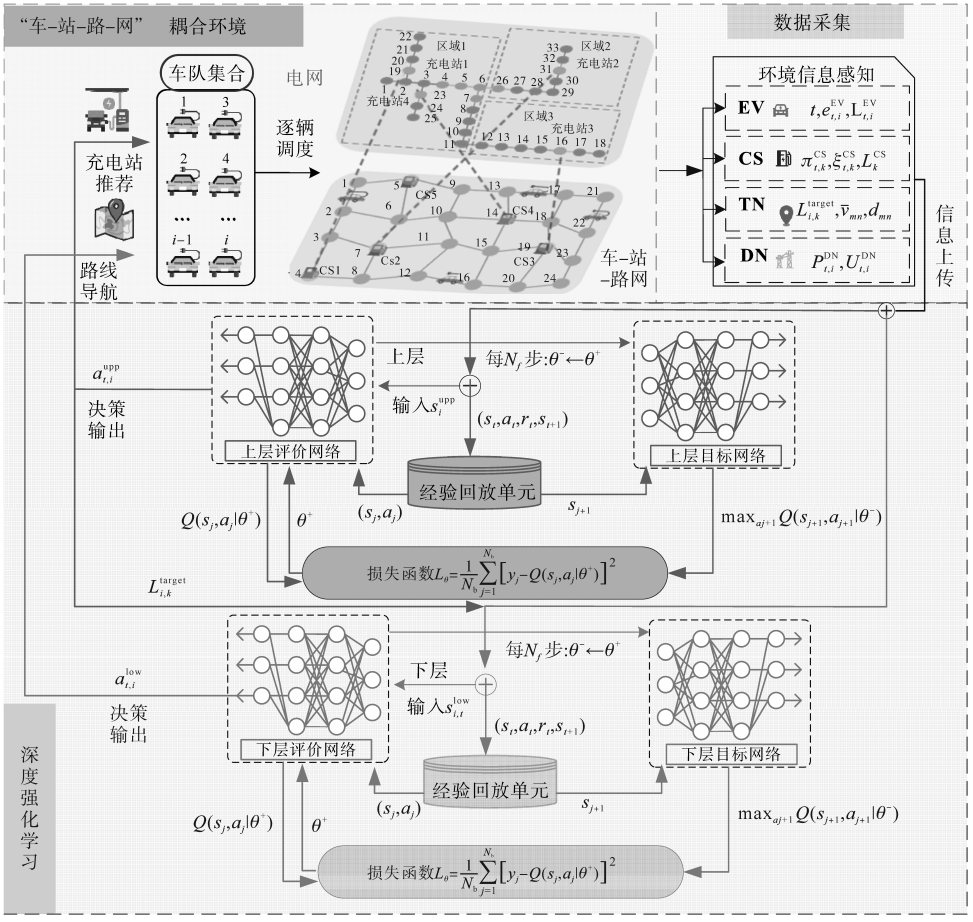

图1　基于HDRL的电动汽车充电引导整体框架

Fig.1　Overall framework of EV charging guidance based on HDRL

2　问题建模

2.1　数学建模

本文从降低EV用户综合成本出发建立多优化目标,综合成本f包括用户费用成本f_1以及时间成本f_2两方面,具体计算公式如下:

$$f=\min_{\varphi_{i,mn},\varphi_{i,k}} f_1+\min_{\varphi_{i,mn},\varphi_{i,k}} f_2 \tag{1}$$

其中

$$\min_{\varphi_{i,mn},\varphi_{i,k}} f_1=C_i^{en}+C_i^{ch}+C_i^{hu} \tag{2}$$

$$\min_{\varphi_{i,mn},\varphi_{i,k}} f_2=\varpi(T_i^{tr}+T_i^{wt}+T_i^{ch}) \tag{3}$$

$$C_i^{en}=\bar{\pi}^{CS}\mu\sum_{\beta_{mn}\in\Omega_i} d_{mn}\varphi_{i,mn} \tag{4}$$

$$C_i^{ch}=\sum_{t=t_i^{sta}}^{t_i^{end}} \pi_{k,t}^{CS}P^{ch}\Delta t\,\varphi_{i,k} \tag{5}$$

$$T_i^{tr}=\sum_{\beta_{mn}\in\Omega_i}\frac{d_{mn}}{\bar{v}_{mn}}\varphi_{i,mn} \tag{6}$$

$$T_i^{\mathrm{ch}}=\frac{Q_i\left(e_i^{\mathrm{exp}}-e_i^{\mathrm{arr}}\right)}{P^{\mathrm{ch}}\eta^{\mathrm{ch}}}\varphi_{i,k} \tag{7}$$

式中:C_i^{en},C_i^{ch}分别为路程耗电费用以及用户充电费用;T_i^{tr},T_i^{wt},T_i^{ch}分别为第i个用户的驾驶时长、充电等待时长以及充电时长,$i\in\Omega^{\mathrm{EV}}$,Ω^{EV}为电动汽车数量集合;$\varphi_{i,mn}$为路径选择变量,$\varphi_{i,mn}=1$表示第i个用户选择交通节点β_m和β_n之间的道路的β_{mn},否则$\varphi_{i,mn}=0$;$\beta_m,\beta_n\in\Omega^{\mathrm{E}}$,$\Omega^{\mathrm{E}}$为交通路网$G^{\mathrm{TN}}$路段集合;$\Omega_i$为第$i$个用户的路径选择集合;$\varphi_{i,k}$为充电站选择变量,$\varphi_{i,k}=1$表示第$i$个用户被推荐至第$k$座充电站,否则$\varphi_{i,k}=0$;$k\in\Omega^{\mathrm{CS}}$,$\Omega^{\mathrm{CS}}$为充电站的集合;$\varpi$为时间成本系数;$\bar{\pi}^{\mathrm{CS}}$为充电站平均充电价格;$\mu$为EV单位能耗;$d_{mn}$为交通路网$G^{\mathrm{TN}}$的道路$\beta_{mn}$的长度;$\pi_{k,t}^{\mathrm{CS}}$为充电站$k$的$t$时刻的实时电价,$t\in T$,$T$为控制时间;$P^{\mathrm{ch}}$为充电功率;$t_i^{\mathrm{sta}}$,$t_i^{\mathrm{end}}$分别为第$i$个用户的充电开始时间与充电结束时间;$\bar{v}_{mn}$为交通路网$G^{\mathrm{TN}}$的道路$\beta_{mn}$平均通行速度;$Q_i$为第$i$个用户的电池容量;$e_i^{\mathrm{arr}}$,$e_i^{\mathrm{exp}}$分别为第$i$个用户到达时刻荷电状态(state of charge,SOC)以及期望结束的SOC;η^{ch}为充电设备效率。

1)电动汽车约束如下:

$$e_i^{\mathrm{req}}-\frac{\mu\sum_{\beta_{mn}\in\Omega_i}d_{mn}}{Q_i^{\mathrm{bat}}}\varphi_{i,mn}>e^{\mathrm{fl}} \tag{8}$$

$$\sum_{k\in\Omega^{\mathrm{CS}}}\varphi_{i,k}=1 \tag{9}$$

$$\sum_{\beta_{mn}\in\Omega^{\mathrm{E}}}\varphi_{i,mn}=1 \tag{10}$$

式中:e_i^{req}为充电需求时刻的SOC;e^{fl}为车辆电池最低SOC,低于该值则认为EV抛锚。

约束式(8)限制了车辆的能量范围,式(9)与式(10)则分别约束了充电站选择与路径选择。

2)配电网约束。在电网运行过程中,维持网络节点的电压在一个合理可控的范围是一项重要任务。而规模化的聚集充电行为将会导致节点负荷急剧增大,节点电压受此影响会发生跌落。因此,本文所建立的模型中必须考虑配网的潮流约束和安全约束,如下式所示:

$$-P_{t,i}^{\mathrm{CS}}-P_{t,i}^{\mathrm{load}}=U_{t,i}\sum_{j\in i}U_{t,j}\left(G_{ij}\cos\theta_{t,ij}+B_{ij}\sin\theta_{t,ij}\right) \tag{11}$$

$$-Q_{t,i}^{\mathrm{CS}}-Q_{t,i}^{\mathrm{LOAD}}=U_{t,i}\sum_{j\in i}U_{t,j}\left(G_{ij}\sin\theta_{t,ij}-B_{ij}\cos\theta_{t,ij}\right) \tag{12}$$

$$U_i^{\min}\leqslant U_{t,i}\leqslant U_i^{\max} \tag{13}$$

$$I_{ij}^{\min}\leqslant I_{ij,t}\leqslant I_{ij}^{\max} \tag{14}$$

式中:$P_{t,i}^{\mathrm{CS}}$,$Q_{t,i}^{\mathrm{CS}}$分别为配电网节点v_i的充电有功负荷与无功负荷;$U_{t,i}$,$U_{t,j}$分别为配网节点v_i和配网节点v_j的实时电压,$v_i,v_j\in\Omega^{\mathrm{DN}}$,$\Omega^{\mathrm{DN}}$为配电网节点集合;$P_{t,i}^{\mathrm{load}}$,$Q_{t,i}^{\mathrm{CS}}$分别为常规有功与无功负荷;$G_{ij}$,$B_{ij}$分别为支路电导与电纳;$\theta_{t,ij}$为相角差;$U_i^{\max}$,$U_i^{\min}$分别为节点电压上、下限;$I_{ij}^{\max}$,$I_{ij}^{\min}$分别为电流上、下限。

本文以基准电压10.6 kV的拓扑作为仿真配网环境[20],因此电压上、下限分别设置为0.95(标幺值)与1.05(标幺值)。

3)交通路网约束。交通路网模型为研究路径规划的基础,因此,引入图论分析方法对城市交通路网G^{TN}进行建模描述。针对给定的交通路网G^{TN}进行如下建模[21]:

$$\begin{cases}G^{\mathrm{TN}}=(\Omega^{\mathrm{N}},\Omega^{\mathrm{E}},\Omega^{\mathrm{W}})\\ \Omega^{\mathrm{N}}=\left\{\beta_m\middle|m\in\Omega^{\mathrm{N}}\right\}\\ \Omega^{\mathrm{E}}=\left\{\beta_{mn}\middle|\beta_m\in\Omega^{\mathrm{N}},\beta_n\in\Omega^{\mathrm{N}},m\neq n\right\}\\ \Omega^{\mathrm{W}}=\left\{w_{mn}\middle|\beta_{mn}\in\Omega^{\mathrm{E}}\right\}\end{cases} \tag{15}$$

式中:Ω^{N}为节点集合,即交通路网G^{TN}的节点集合;Ω^{E}为有向弧段的集合,即交通路网G^{TN}的路段集合;Ω^{W}为路段权值集合,即交通路网G^{TN}的道路路阻,表示路段的量化属性。

其中路段长度、通行速度、行程时间以及出行费用等可作为路段权值w_{mn}进行量化研究。

进一步,为对交通路网G^{TN}进行量化赋值,通过邻边矩阵$\boldsymbol{E}=a_{mn}$将道路路阻分配到各路段上。

$$a_{mn}=\begin{cases}w_{mn} & \beta_{mn}\in\Omega^{\mathrm{E}}\\ 0 & \beta_m=\beta_n\\ \infty & \beta_{mn}\notin\Omega^{\mathrm{E}}\end{cases} \tag{16}$$

邻边矩阵$\boldsymbol{E}$最终表示如下:

$$\boldsymbol{E}=\begin{bmatrix}0 & w_{12} & w_{13} & \cdots & \infty\\ w_{21} & 0 & w_{23} & \cdots & \infty\\ \infty & w_{32} & 0 & \cdots & \infty\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \infty & \infty & \infty & \cdots & 0\end{bmatrix} \tag{17}$$

式中:∞为节点β_m与β_n之间不存在连接路段。

因此,电动汽车车主通过搜索路段权值w_{mn}进行最优路径规划。行驶路线的拓扑限制表示如下:

$$\sum_{n=L_{t,i}^{\mathrm{EV}}}^{L_k^{\mathrm{CS}}}\varphi_{i,mn}-\sum_{n=L_{t,i}^{\mathrm{EV}}}^{L_k^{\mathrm{CS}}}\varphi_{i,mn}=\begin{cases}1 & m=L_{t,i}^{\mathrm{EV}}\\ 0 & m\neq L_{t,i}^{\mathrm{EV}},L_k^{\mathrm{CS}}\\ -1 & m=L_k^{\mathrm{CS}}\end{cases} \tag{18}$$

式中:$L_{t,i}^{\mathrm{EV}}$为当前位置;L_k^{CS}为分配充电站位置。

电动汽车从$L_{t,i}^{\mathrm{EV}}$到L_k^{CS}的路径按照起始节点、中间节点以及终止节点进行规划,保证所选路线可以顺序连接。

2.2　基于双层FMDP的电动汽车充电与路径决策构建模型

强化学习方法通过智能体与环境的交互,采用探索-利用策略进行试错,以获得奖励值,并基于此建立状态与动作的最优映射关系。在电动汽车充电引导问题中,车主作为智能体,感知交通和电气化环境信息,并通过获得的奖励值做出充电站选择和行驶路线的决策,直至到达目的地。这一过程符合有限马尔科夫链的决策模型,即有限马尔科夫决策过程,它被广泛应用于包括配电网控制和能量调度在内的时序决策问题。针对电动汽车充电引导问题,本文提出双层FMDP模型,解耦充电站推荐与路径选择,以实现更高效的调度。本文所提模型的具体建模如下文所示。

2.2.1　状态

状态代表智能体对环境信息的实时感知。状态空间则代表所有可能状态的集合。

1)上层网络。考虑到充电站的推荐本质是EV充电需求与CS能量资源的时空匹配问题,将上层状态s_i^{upp}分为EV,CS与DN三方面的实时信息,即

$$s_i^{\mathrm{upp}}=\{\underbrace{t,e_{t,i}^{\mathrm{EV}},L_{t,i}^{\mathrm{EV}}}_{\mathrm{EV}},\underbrace{\pi_{t,k}^{\mathrm{CS}},\xi_{t,k}^{\mathrm{CS}},L_k^{\mathrm{CS}}}_{\mathrm{CS}},\underbrace{P_{t,i}^{\mathrm{DN}},U_{t,i}^{\mathrm{DN}}}_{\mathrm{DN}}\} \tag{19}$$

式中:$e_{t,i}^{\mathrm{EV}}$为EV的t时刻的实时SOC,在上层中该值即为$e_{t,i}^{req}$;$\xi_{t,k}^{\mathrm{CS}}$为充电站状态变量,$\xi_{t,k}^{\mathrm{CS}}\geqslant 0$表示站内空闲桩数量,否则表示排队等待人数;$P_{t,i}^{\mathrm{DN}}$,$U_{t,i}^{\mathrm{DN}}$分别为配电网节点的负荷与电压。

2)下层网络。一旦上层输出了目标充电站,下层将向着该目的地进行导航。因此,下层状态$s_{i,t}^{\mathrm{low}}$可以表示为

$$s_{i,t}^{\mathrm{low}}=\{\underbrace{t,e_{t,i}^{\mathrm{EV}},L_{t,i}^{\mathrm{EV}}}_{\mathrm{EV}},\underbrace{L_{i,k}^{\mathrm{target}},\bar{v}_{mn},d_{mn}}_{\mathrm{TN}}\} \tag{20}$$

式中:$L_{i,k}^{\mathrm{target}}$为目标充电站位置。

2.2.2　动作

动作是在给定环境下智能体所做出的决策。

1)上层网络。上层动作输出被定义为所推荐充电站的索引,其数学描述如下式所示:

$$a_i^{\mathrm{upp}}=\{L_k^{\mathrm{CS}}\} \tag{21}$$

2)下层网络。对于一个给定路网G^{TN},驾驶导航问题是一个离散的路径选择问题,其正是下层所需要解决的。下层的动作决策可以表示为下式:

$$a_{t,i}^{\mathrm{low}}=\{\beta_{mn}\} \tag{22}$$

将得到的一系列决策动作$a_{t,i}^{\mathrm{low}}$依次连接构建最优导航路径$\Psi_{t,i}^{\mathrm{low}}$。EV根据该路径$\Psi_{t,i}^{\mathrm{low}}$从当前位置$L_{i,k}^{\mathrm{EV}}$进行行驶直至抵达目标充电站的位置$L_{i,k}^{\mathrm{target}}$。最优导航路径$\Psi_{t,i}^{\mathrm{low}}$构建如下:

$$\Psi_{t,i}^{\mathrm{low}}=\sum_{L_{t,i}^{\mathrm{EV}}}^{L_{i,k}^{\mathrm{target}}}a_{t,i}^{\mathrm{low}} \tag{23}$$

2.2.3　奖励

智能体在执行动作后的及时反馈,是帮助智能体学习特定能力的重要一环。

1)上层网络。在上层,充电站的动作选择将会直接影响用户到站后的充电费用C_i^{ch}、等待时间T_i^{wt}以及充电时间T_i^{ch}。与此同时,规模化EV的充电选择则会影响到配电网的运行状态。因此,将这三项作为第一层智能体的奖励函数,如下式所示:

$$r_i^{\mathrm{upp}}=-C_i^{\mathrm{ch}}-\omega(T_i^{\mathrm{wt}}+T_i^{\mathrm{ch}})-\frac{1}{N^{\mathrm{DN}}}\sum_{i\in\Omega^{\mathrm{DN}}}\left|\frac{U_{t,i}-U_i^*}{U_i^*}\right| \tag{24}$$

式中:U_i^*为配电网节点的额定电压;N^{DN}为配网节点的数量。

2)下层网络。结合总体优化目标以及上层奖励设计,下层智能体的奖励主要包括用户导航过程中的电池电量成本以及道路时间成本。一旦车辆抵达目标充电站,将给予智能体一个正的奖励。相反,如果EV在电池耗尽前仍没有抵达目的地,用户则需要自行呼叫拖车救援。因此,智能体将会得到一个负的惩罚。为此,将下层奖励值作为交通路网G^{TN}的道路路阻w_{mn},并且赋值到各交通路段β_{mn}上。智能体通过对比各道路路阻w_{mn}带来的奖励或者惩罚的反馈确定所要采取的当前动作,如下式所示:

$$w_{mn}=r_{t,i}^{\mathrm{low}}=\begin{cases}-d_{mn}\varphi_{mn}\bar{\pi}^{\mathrm{CS}}\mu+\dfrac{\omega d_{mn}}{\bar{v}_{mn}}\varphi_{i,mn} & L_{t+1,i}^{\mathrm{EV}}\neq L_{i,k}^{\mathrm{target}}\\ \omega^{\mathrm{arr}} & L_{t+1,i}^{\mathrm{EV}}=L_{i,k}^{\mathrm{target}}\\ -\omega^{\mathrm{tow}} & e_{t,i}^{\mathrm{EV}}-\dfrac{\mu d_{mn}}{Q_i}\varphi_{i,mn}<e^{\mathrm{flat}}\end{cases} \tag{25}$$

式中:$L_{t+1,i}^{\mathrm{EV}}$为EV下一时刻位置;ω^{arr}为一个显著的导航成功奖励;$-\omega^{\mathrm{tow}}$为导航失败惩罚项,即该区域的拖车成本。

EV下一时刻位置 $L_{t+1,i}^{\mathrm{EV}}$ 由当前位置 $L_{t,i}^{\mathrm{EV}}$ 以及路径选择动作 $a_{t,i}^{\mathrm{low}}$ 决定。

2.2.4　状态-动作价值函数

状态-动作价值函数用于评价基于当前策略 π 下执行动作后,智能体能够获得的累计期望奖励。虽然上层与下层智能体分别依赖策略 π^{upp} 与 π^{low},它们的状态-动作价值函数 $Q^{\Psi}(s,a)$(Q-value)相同,如下式所示:

$$Q^{\Psi}(s,a)=\mathbb{E}\left[\sum_{h=0}^{H}\gamma^{h}r_{t+h}\middle|s_t=s,a_t=a\right] \tag{26}$$

式中:H 为时间步的视野;γ 为折扣率。

在EV充电导航问题中,智能体的目标是求取最优策略 Ψ^{*},其等价于求取能够获得最大 $Q^{\Psi}(s,a)$ 的策略:

$$Q^{\Psi^{*}}(s,a)=\max_{\Psi}Q^{\Psi}(s,a) \tag{27}$$

3　基于Rainbow架构的求解方法

本文提出的求解方法采用Rainbow算法,该算法基于DQN架构,通过整合双重DQN机制、优先级经验回放、Dueling网络结构、辍学层技术和学习率衰减等机制,有效提升了算法在双层决策问题求解中的学习效能、决策精准度和探索能力,同时增强了模型的泛化性和适应性。因此,本文基于Rainbow算法进行双层决策问题的求解。

本文所提方法的训练流程如图2所示。单次回合中,逐辆进行推荐电动汽车,上层智能体获取上层环境状态 $s_{i,k}^{\mathrm{upp}}$,并基于上层评价网络选取目标充电站,然后将该目标充电站 a_i^{upp} 传送至下层智能体的状态空间 $s_{i,t}^{\mathrm{low}}$。

其次,下层智能体接收到目标充电站位置 $L_{i,k}^{\mathrm{target}}$ 时,开始获取下层环境状态 $s_{i,t}^{\mathrm{low}}$,执行动作 $a_{t,i}^{\mathrm{low}}$,计算智能体所获得的即时奖励 $r_{t,i}^{\mathrm{low}}$,观察新的环境状态 $s_{i,t+1}^{\mathrm{low}}$,并将样本 $(s_{i,t}^{\mathrm{low}},a_{t,i}^{\mathrm{low}},r_{t,i}^{\mathrm{low}},s_{t,i+1}^{\mathrm{low}})$ 存放至经验池 D^{low} 中。通过上述交互累积历史样本,根据TD-error和优先回放机制,抽取高误差样本组成mini-batch。接着,对下层评价网络进行梯度下降,更新网络参数 $\theta^{\mathrm{low},+}$,间隔 N_r 步将其复制到下层目标网络 $\theta^{\mathrm{low},-}$。

最后,当车主到达后,计算上层智能体对应奖励 r_i^{upp} 和下一状态 $s_{i,k+1}^{\mathrm{upp}}$,并按照与下层智能体相同的更新机制来优化其网络。特别地,学习率 α_n 将随着回合数倒数衰减。重复上述步骤直至达到预设的最大回合数。

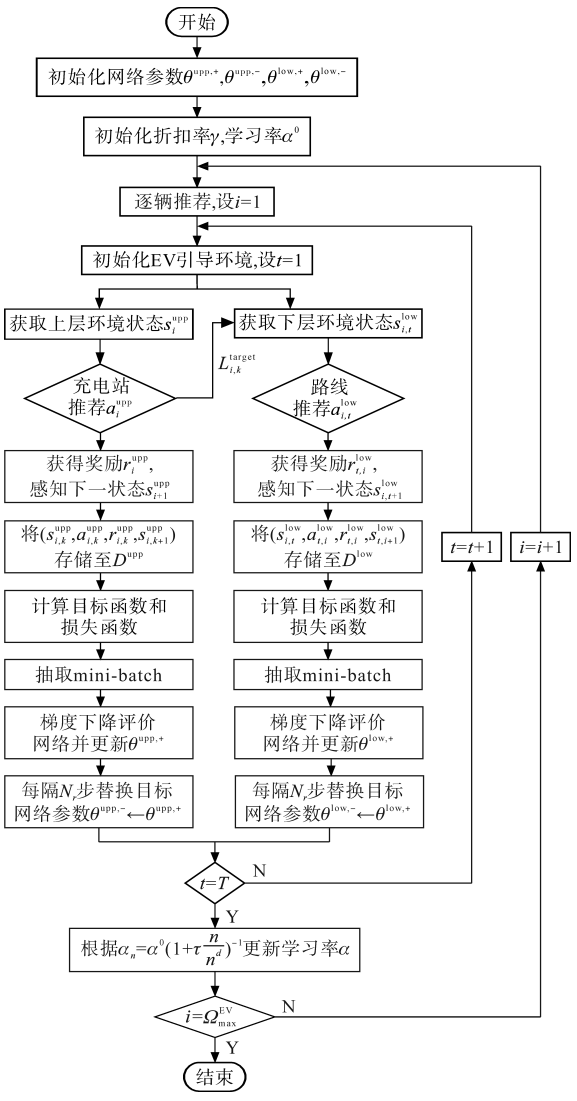

图2　基于Rainbow算法的求解流程

Fig.2　Solution process based on the Rainbow algorithm

4　算例分析

4.1　算例配置

本文采用南京市实际路网拓扑和运营充/换电站构建仿真环境,如图3所示。其中,本次实验中使用的城市道路实时运行通行数据来自Open Street Map平台(https://www.openstreetmap.org/),同时使用的充/换电站的配置和运营数据来自充电吧平台(http://admin.bjev520.com/jsp/beiqi/pc-map/do)。为了和实际交通路网进行匹配,采用IEEE 33节点配电网构建城市电网,且10座充电站分别接入于节点4,6,9,13,16,17,20,24,28和节点32。本文初始参数配置如表1所示。训练环境采用CPU I9 9960X,GPU RTX2070,RAM 32GB。

表1　仿真环境初始参数配置

Tab.1　Initial parameter configuration for the simulation environment

参数名称	参数配置
仿真车辆数	1 500 辆
EV电池容量	40 kW·h
时间成本系数ϖ	86.02 元/h
单位距离能耗μ	0.2 kW·h/km
导航成功奖励w^{arr}	100
导航失败惩罚w^{tow}	200
最低SOCe^{flat}	0.05

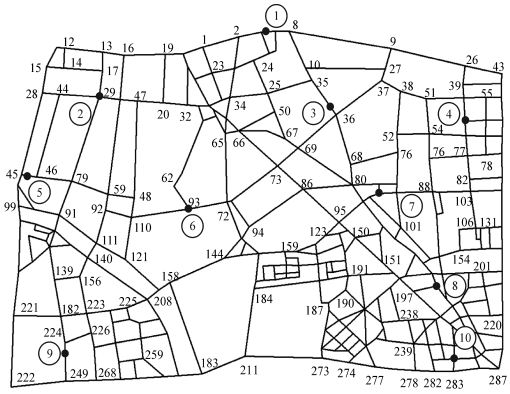

图3　路网拓扑和运营的充/换电站位置

Fig.3　Road network topology and location of operated charging/swapping stations

4.2　训练过程

图4给出了本文所提HDRL策略基于Rainbow架构算法求解得到的上、下两层训练奖励曲线。由图4可知,上、下两层智能体的奖励初期均为探索发散过程,经过一段时间训练后才达到收敛,而收敛区间各不相同。上层智能体大约在训练300回合收敛于最大奖励值-86元。然而,下层智能体需要多训练500回合才能趋于稳定,并稳定在10元附近。二者前期都呈现不稳定的趋势,因为智能体对环境进行较大程度的探索,并不断修正策略,直到探寻到最优的充电站和充电行驶路线。由于下层智能体所在路网环境更加复杂,且需要考虑到上层的充电站选择策略,则下层智能体需要花费更多的回合来探索最优路线引导策略。

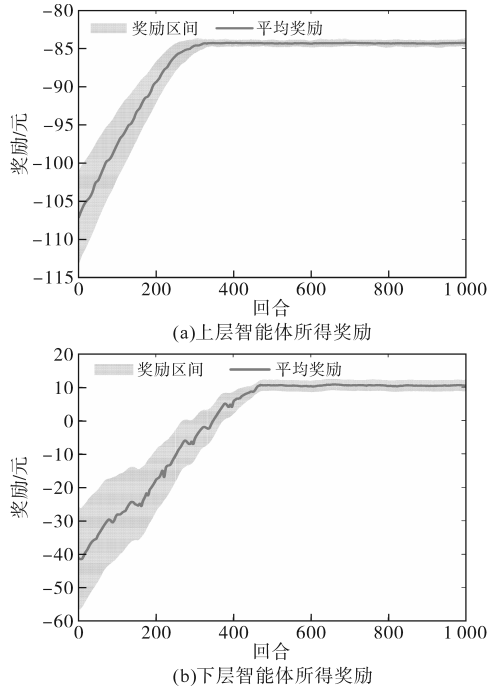

图4　基于Rainbow架构方法的上下层的训练奖励

Fig.4　Training rewards for the upper and lower layers based on the Rainbow architecture method

4.3　测试结果

图5为本文所提基于HDRL的方法和基于无序引导方法的引导结果对比图。基于无序引导方法是上层决策直接选择离充电触发位置距离最近的充电站。而下层决策通过Dijkstra算法选择最短的行驶路线。

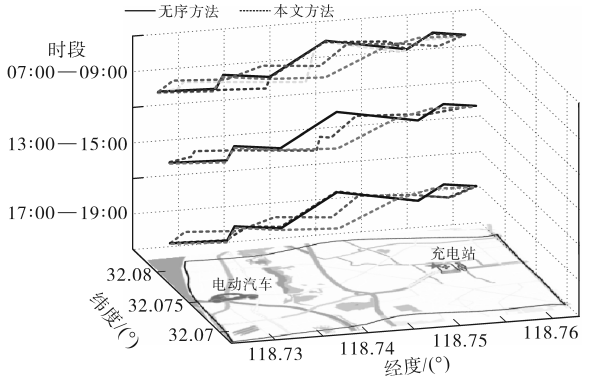

图5　不同时段下HDRL与无序引导方法的路径规划结果

Fig.5　Path planning results of HDRL and disorder guidance methods at different time periods

由图5可知,电动汽车在三个不同的时间段进行测试,都从相同的出发点行驶向一个相同的充电站。早上选择07:00—09:00时间段,下午选择13:00—15:00时间段,晚上选择17:00—19:00时间段,每个时间段内无序决策引导下只有单条路线可供选择。而经过本文所提方法引导,一共规划七条充电行驶路径,EV用户可选择不同的路线,有效地避免了交通拥堵等意外事件,提高了出行效率和便捷性。此外,通过对比三个时间段的引导路线可知,三个时间段内除了无序方法引导的路线相同外,其它路线存在差异,说明根据不同的时段,不同的交通流量,车主可以选择不同的路线,满足了充电汽车引导的实时性要求。

另外,图6为随时间变化下HDRL与无序引

导方法的电动汽车分布结果。充电站的位置分布具有其特殊性,从整体上看,充电站5~8分布在市区,而其它充电站集中在郊区区域。由图6可知,在充电高峰期11:00—13:00和17:00—21:00,经过无序引导方法引导的电动汽车主要集中于充电站7~8,这容易造成充电拥堵和排队过长的现象,大大影响了充电效率,而经过本文所提方法的引导,EV分布比较分散,充电站4~8都存在适量的充电汽车,有效缓解了高峰期的充电压力。

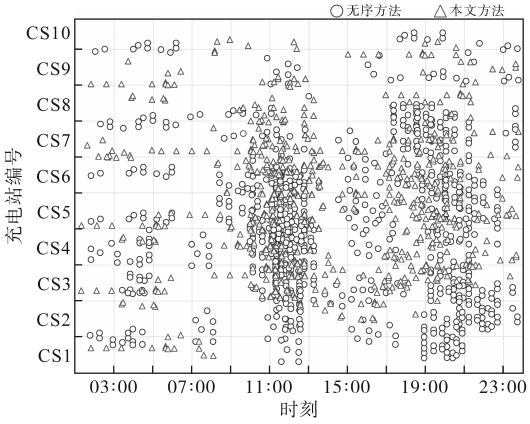

图6　不同时间下HDRL与无序引导方法的EV分布结果

Fig.6　EV distribution results of HDRL and disorder guidance methods at different time periods

图7为不同时间下HDRL与无序引导方法的总时间花费和总费用花费对比图。表2为HDRL与无序引导方法的时间和成本花费平均值。由图表可知,在无序方法引导下,09:00—10:00时段到达第一个总时间峰值区间,20:00—22:00时段到达第二个总时间峰值区间,用户的平均总时间花费为61.12 min。在13:00时附近用户花费的总费用达到了最大值52.26元,一天平均费用花费约为30.15元。本文所提方法在成本方面均有所降低,车主在09:00—10:00期间平均时间花费仅为56.55 min,降低了6.4%,同时在20:00—22:00期间成本为54.57 min,降低了10.72%。在13:00时左右花费的成本降低了5.08元,一天平均总费用为24.67元,降低了18.18%。

相比无序引导方法,本文方法综合考虑了电动汽车用户、充电站、交通网和配电网的利益,通过对存在充电需求的车辆进行有序引导,实时输出最优充电站位置和充电行驶路线,避免了大量车辆集中涌向少数热门充电站而导致排队等待时间过长的问题,车主能更快地抵达充电站或者目的地,有效缩短了整体的时间花费,同时也减

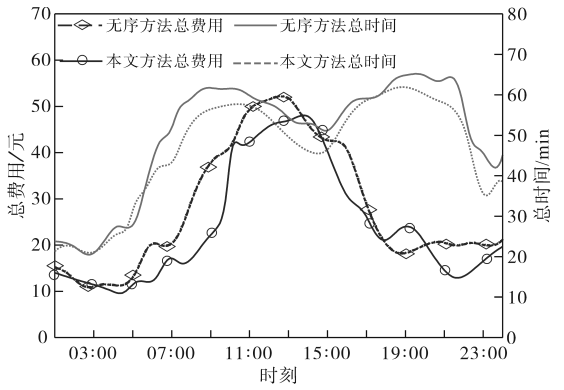

图7　不同时间下HDRL与无序引导方法的时间和成本花费结果

Fig.7　Time and cost of HDRL and disorder guidance methods at different times

表2　HDRL与无序引导方法的时间和成本花费平均值

Tab.2　Average time and cost of HDRL versus disorder guidance methods

时间	无序方法总费用/元	本文方法总费用/元	无序方法总时间/min	本文方法总时间/min
平均值	30.15	24.67	51.32	45.49

少了行车成本。

进一步,图8展示了在HDRL与无序引导方法引导下在20:00时刻的配电网节点电压分布。由图8可知,在配电网的13~17节点和30~33节点,无序方法下都出现了电压违规现象,电压值低于规定的低电压阈值0.95(标幺值),对电网的稳定性造成了影响。而采用本文所提的方法则电压未发生电压越限现象。说明本文所提HDRL方法可减少电动汽车聚集充电,实现EV的有效充电引导,缓解了配电网压力,提高了充电的安全性和稳定性。

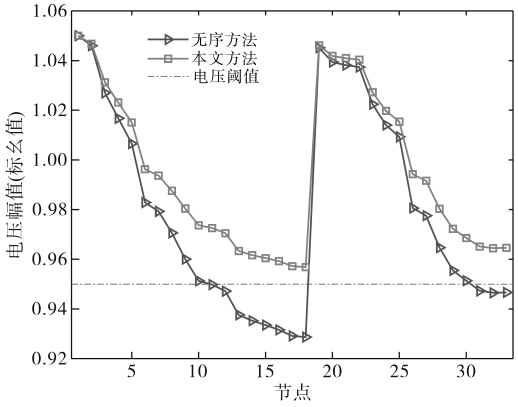

图8　HDRL与无序引导方法在20:00节点电压分布

Fig.8　HDRL and disorder guidance method at 20:00 node voltage distribution

4.4　方法对比

最后,本文选择其他两种DRL策略来综合对比HDRL的应用效果,其中深度Q网络采用传统

FMDP训练范式,竞争深度网络(dueling deep Q network,DDQN)采用本文的基于约束条件转换成奖励惩罚的训练范式。图9和图10分别给出了三种方法持续100 d的累积平均总成本(费用成本和时间成本之和)和在20:00时刻的配电网节点电压分布。由图可知,采用DQN累计的100 d平均总成本最低,约为11 000元,因为基于DQN方法的架构简单且计算效率高。

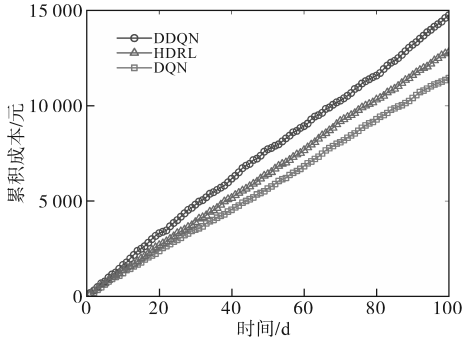

图9　不同策略持续100天的累计平均总成本

Fig.9　Cumulative average total cost of different strategies over a period of 100 days

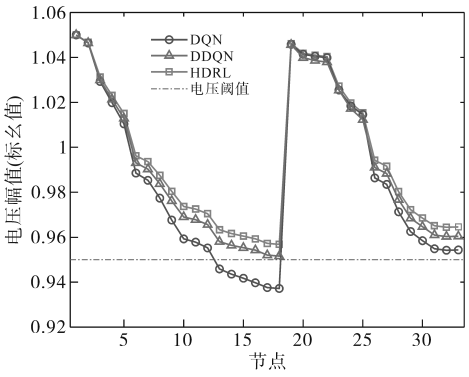

图10　不同策略下节点电压分布

Fig.10　Node voltage distribution under different strategies

然而传统的DQN中的奖励没有考虑到安全约束,DQN的训练过程主要侧重于最大化累积奖励,而没有将安全约束(如电压越限)直接纳入优化目标中,造成了电压越限现象。虽然DDQN和HDRL方法的平均总成本均高于DQN方法,分别为15 000元和12 500元,然而HDRL和DDQN未出现电压越限现象。

此外,表3给出了上述三种方法的评估指标,三种方法均能在0.17 s完成决策制定。由于HDRL是基于Rainbow架构,引入了多种改进机制,智能体经过训练得到的累计奖励比较低,HDRL的总费用和总时间均低于DDQN。但更多的改进机制所需的训练时间也有所增加,分别比DQN和DDQN方法多出87.9%和57.4%。

表3　DQN,DDQN和HDRL策略的评价指标对比

Tab.3　Comparison of the evaluation indicators for DQN,DDQN and HDRL strategies

方法	总费用/元	总时间/元	训练时间/h	决策时间/s
DQN	49.24	57.26	5.31	0.16
DDQN	57.89	67.59	6.34	0.16
HDRL	54.12	61.23	9.98	0.17

5　结论

本文提出了基于分层深度强化学习的实时充/换电引导策略。通过构建双层决策架构,有效地将充电站推荐问题和充电行驶路径问题解耦,并基于Rainbow算法架构融合多种网络改进机制,对所提问题进行求解,最后在一个真实的路网环境中进行了仿真验证。实验结果显示本文所提方法能够帮助EV车主节省充电成本和充电时间,与无序引导方法相比,一天平均总费用能够降低20%。此外,在缓解充电站充电压力和提高配电网稳定性方面也具有优势。

尽管在DQN方法框架内加入安全约束理论上是可行的,并且能够提升智能体在复杂电网环境中的决策安全性,但目前的AI方法往往未能充分考虑安全约束对决策结果的影响。未来的工作将深入研究如何将安全约束有效地融入DQN算法中,通过优化网络结构和调整训练参数,使智能体能够在满足安全约束的前提下做出最优决策。

参考文献

[1]　黄学良,刘永东,沈斐,等. 电动汽车与电网互动:综述与展望[J]. 电力系统自动化,2024,48(7):3-23.
HUANG Xueliang,LIU Yongdong,SHEN Fei,et al. Vehicle to grid:review and prospect[J]. Automation of Electric Power Systems,2024,48(7):3-23.

[2]　李恒杰,夏强强,史一炜,等. 考虑目标充电站选择冲突的电动汽车充电引导策略[J]. 电力自动化设备,2022,42(5):68-74.
LI Hengjie,XIA Qiangqiang,SHI Yiwei,et al. Electric vehicle charging guidance strategy considering selection conflict of target charging stations[J]. Electric Power Automation Equipment,2022,42(5):68-74.

[3]　玉少华,杜兆斌,陈丽丹,等. 融合路网-电网信息的电动汽车充放电行为引导与调控策略[J]. 电力系统自动化,2024,48(7):169-180.
YU Shaohua,DU Zhaobin,CHEN Lidan,et al. Guidance and regulation strategy for charging and discharging behaviors of electric vehicles based on fusion of road network and power

grid information[J]. Automation of Electric Power Systems, 2024,48(7):169-180.

[4] 邵尹池,穆云飞,林佳颖,等."车-站-网"多元需求下的电动汽车快速充电引导策略[J].电力系统自动化,2019,43(18):60-68.
SHAO Yinchi, MU Yunfei, LIN Jiaying, et al. Fast charging guidance strategy for multiple demands of electric vehicle, fast charging station and distribution network[J]. Automation of Electric Power Systems,2019,43(18):60-68.

[5] 刘珏,穆云飞,董晓红,等.考虑交通时滞的配电-交通耦合系统协同优化运行策略[J].电力系统及其自动化学报,2024,36(10):49-59.
LIU Jue, MU Yunfei, DONG Xiaohong, et al. Collaborative operation strategy for power-traffic coupling system considering traffic delay[J]. Proceedings of the CSU-EPSA, 2024, 36(10): 49-59.

[6] LU H, SHAO C, HU B, et al. En-route electric vehicles charging navigation considering the traffic-flow-dependent energy consumption[J]. IEEE Transactions on Industrial Informatics, 2022,18(11):8160-8172.

[7] 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. Route planning and charging navigation strategy for electric vehicles based on real-time traffic information[J]. Proceedings of the CSEE, 2020, 40 (2): 534-550.

[8] 袁晓冬,甘海庆,王明深,等.车联网环境下电动汽车主动充电引导模型[J].电力系统自动化,2024,48(7):159-168.
YUAN Xiaodong, GAN Haiqing, WANG Mingshen, et al. Active charging guidance model of electric vehicles based on internet of vehicles[J]. Automation of Electric Power Systems, 2024,48(7):159-168.

[9] 苏粟,王建祥,王磊,等.基于动态哈夫模型及双边匹配的电动汽车充电引导策略[J].电力系统自动化,2024,48(7):181-189.
SU Su, WANG Jianxiang, WANG Lei, et al. Guidance strategy for electric vehicle charging based on dynamic huff model and bilateral matching[J]. Automation of Electric Power Systems, 2024,48(7):181-189.

[10] TAO Y, QIU J, LAI S, et al. Distributed electric vehicle assignment and charging navigation in cyber-physical systems[J]. IEEE Transactions on Smart Grid, 2023, 15(2): 1861-1875.

[11] QIAN T, SHAO C, WANG X, et al. Deep reinforcement learning for EV charging navigation by coordinating smart grid and intelligent transportation system[J]. IEEE Transactions on Smart Grid, 2019, 11(2): 1714-1723.

[12] ZHANG Z, CHEN Z, GÜMRÜKCÜ E, et al. A nudge-based approach for day-ahead optimal scheduling of destination charging station with flexible regulation capacity[J]. IEEE Transactions on Transportation Electrification, 2024, 10 (4): 8498-8512.

[13] RAN L, QIN J, WAN Y, et al. Fast charging navigation strategy of EVs in power-transportation networks: a coupled network weighted pricing perspective[J]. IEEE Transactions on Smart Grid, 2024, 15(4): 3864-3875.

[14] 张文昕,栗然,臧向迪,等.基于强化学习的电动汽车换电站实时调度策略优化[J].电力自动化设备,2022,42(10):134-141.
ZHANG Wenxin, LI Ran, ZANG Xiangdi, et al. Real-time scheduling strategy optimization for electric vehicle battery swapping station based on reinforcement learning[J]. Electric Power Automation Equipment, 2022, 42 (10): 134-141.

[15] SU S, LI Y, YAMASHITA K, et al. Electric vehicle charging guidance strategy considering "traffic network-charging station-driver" modeling: a multi-agent deep reinforcement learning based approach[J]. IEEE Transactions on Transportation Electrification, 2023, 10(3): 4653-4666.

[16] LI Y, SU S, ZHANG M, et al. Multi-agent graph reinforcement learning method for electric vehicle on-route charging guidance in coupled transportation electrification[J]. IEEE Transactions on Sustainable Energy, 2023, 15(2): 1180-1193.

[17] 江昌旭,卢玥君,邵振国,等.基于图神经网络多智能体强化学习的电力-交通融合网协同优化运[J].高电压技术,2023,49(11):4622-4631.
JIANG Changxu, LU Yuejun, SHAO Zhenguo, et al. Collaborative optimization operation of integrated electric power and traffic network based on graph neural network multi-agent reinforcement learning[J]. High Voltage Engineering, 2023, 49 (11): 4622-4631.

[18] 曹昉,胡佳彤,罗进奔,等.基于路网动态模型下EV路径模拟的快速充电站容量配置[J].电力自动化设备,2022,42(10):107-115.
CAO Fang, HU Jiatong, LUO Jinben, et al. Capacity configuration of fast charging stations based on EV path simulation under dynamic model of transportation network[J]. Electric Power Automation Equipment, 2022, 42(10): 107-115.

[19] 佟晶晶,温俊强,王丹,等.基于分时电价的电动汽车多目标优化充电策略[J].电力系统保护与控制,2016,44(1):17-23.
TONG Jingjing, WEN Junqiang, WANG Dan, et al. Multi-objective optimization charging strategy for plug-in electric vehicles based on time-of-use price[J]. Power System Protection and Control, 2016, 44 (1): 17-23.

[20] ZHOU Y, YUAN Q, TANG Y, et al. Charging decision optimization for electric vehicles based on traffic-grid coupling networks[J]. Power System Technology, 2021, 45(9): 3563-3570.

[21] ALQAHTANI M, HU M. Dynamic energy scheduling and routing of multiple electric vehicles using deep reinforcement learning[J]. Energy, 2022, 244(Apr. 1 Pt. A): 122626.

收稿日期:2024-11-04
修改稿日期:2024-12-30