DOI:10.12158/j.2096-3203.2021.05.003

基于深度强化学习的充光储能源站调度策略

孙广明[1], 陈良亮[1], 王瑞升[2], 陈中[2], 邢强[2]
(1. 南瑞集团(国网电力科学研究院)有限公司,江苏 南京 211106;
2. 东南大学电气工程学院,江苏 南京 210096)

摘 要:为了应对大规模电动汽车调度模型求解复杂、算力要求高的问题,机器学习方法在电动汽车充电导航调度中越来越受到关注。针对充光储一体化能源站,文中提出了一种基于深度强化学习(DRL)的充光储能源站调度策略。首先,分析了能源站运行策略与 DRL 基本理论。其次,基于后悔理论刻画用户对不同充电方案时间与费用的心理状态,建立了智能体对"人-车-站"状态环境全感知模型,并引入时变 ε-greedy 策略作为智能体动作选择方法以提高算法收敛速度。最后,结合南京市实际道路与能源站分布设计了多场景算例仿真,结果表明所提方法在考虑用户心理效应的基础上能够有效提高能源站光伏消纳率,为电动汽车充电调度提供了一种新思路。

关键词:电动汽车;充光储能源站;充电调度;深度强化学习;后悔理论;全感知模型
中图分类号:TM71 **文献标志码**:A **文章编号**:2096-3203(2021)05-0017-08

0 引言

面对日益严峻的能源危机与环境污染问题,电动汽车(electric vehicle,EV)作为环境友好型交通工具迎来了发展机遇[1—2]。然而规模化 EV 的随机充电行为会导致负荷峰值增加、电能质量降低等问题,给配电网的安全与经济运行带来了挑战[3—4]。同时,面对规模化电动汽车调度算力要求高、计算复杂的问题,传统优化模型无法满足实时调度需求。因此,研究充光储一体化能源站的区域电动汽车优化调度策略,已成为亟待解决的重要问题。

目前,国内外学者在针对光储能源站的电动汽车调度方面已取得一定成果。考虑光伏发电等可再生能源对优化调度策略的影响,文献[5]以能源站运行成本为优化目标,基于多模态近似动态规划进行求解,在不同定价模型与光伏出力情况下均表现出较强鲁棒性。文献[6]以减少微电网与配电网交换功率以及微电网网络损耗为优化目标,采用序列二次规划算法进行求解。通过对 EV 进行充放电调度使日负荷曲线跟踪发电曲线,并网模式下的网络损耗及离网模式下的所需储能系统容量均得到降低。文献[7]考虑能源站源荷互补特性,提出了一种考虑不确定性风险的能源站多时间尺度调度模型。文献[8—9]考虑光伏出力预测误差等不确定性,建立了以充光储能源站日运行成本最小为目标的充电站日前优化模型,并在此基础上建立实时滚动优化模型。文献[10]以大规模 EV 接入的配电网运行成本最小和负荷曲线方差最小为目标建立 EV 优化调度模型,在保证系统运行成本的同时有效降低了负荷峰谷差。

上述研究均建立单/多目标-多约束优化模型解决 EV 调度问题,但应用在实时调度方面均面临着海量计算的压力,无法满足实时调度的需求。同时,上述研究过度依赖模型,当实际应用中包含模型未考虑的不确定性因素时,模型的优化结果得不到保证,算法的鲁棒性与泛化能力有待改进。随着机器学习算法的逐渐成熟,已有少量学者开展了深度强化学习(deep reinforcement learning,DRL)应用于 EV 充电调度的研究。文献[11]提出一种基于竞争深度 Q 网络的充电控制方法,在含高渗透率分布式电源的系统中能够兼顾配电网的安全运行与用户出行需求。文献[12]考虑 EV 行驶距离限制,以最小化 EV 总充电时间为目标,建立 DRL 模型进行训练求解。文献[13]考虑用户用电需求,将 EV 充放电能量边界作为部分状态空间,建立了以最小化功率波动与充放电费用为目标的实时调度模型。文献[14]考虑电价与用户通勤行为的不确定性,从充电电价中提取特征训练 Q 网络,并采用 Q 值最大化原则执行动作。文献[15]以最小化 EV 用户行驶时间与充电成本为目标,利用最短路径法提取当前环境状态训练智能体。

虽然上述研究理解了 DRL 方法的本质,以用户充放电时间或费用作为目标,将车辆与充电站参数作为环境状态进行求解。然而,作为车辆行驶与充电行为的最终执行者,EV 车主对充电方案的感知

收稿日期:2021-04-12;修回日期:2021-06-20
基金项目:国家电网有限公司科技项目"基于大功率 IGBT 的电动汽车能源站柔性控制和主动安全关键技术研究及应用"

效应尤为重要，影响调度策略的可执行性与适用性。为此，文中提出了一种考虑人类行为心理的能源站 EV 调度方法。基于后悔理论刻画 EV 用户心理状态，建立智能体“人-车-站”全状态环境感知模型。同时，引入时变 ε-greedy 策略作为智能体动作选择方法以提高算法收敛速度。最后结合南京市实际道路与能源站分布设计了多场景算例仿真，验证文中所提策略的有效性与实用性。

1 EV 调度问题构建

充光储一体化能源站[16]结构如图 1 所示，按功能可分为：配电网系统、光伏发电系统、储能系统、AC/DC 模块、DC/DC 模块、充电桩、通信管理机以及能量管理系统。

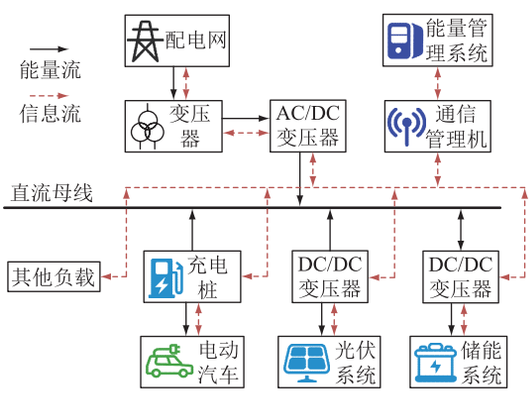


图 1 充光储能源站架构

Fig.1 PV-storage-charging integrated energy station

光伏系统由多组太阳能电池板串并联组成，电池板接收太阳能发电经 DC/DC 变换器接入直流母线，电能主要用于 EV 充电。储能系统由电池组构成，通过双向 DC/DC 变换器接入直流母线。当光伏系统发电有剩余时，其储存电能；当光伏发电不足时，其释放电能。AC/DC 模块为配电网系统与能源站的连接单元，当能源站内部电能不能满足充电需求时由配电网经 AC/DC 接入充电负荷。

针对充光储一体化能源站，考虑能源站光伏消纳能力与 EV 用户利益，可以建立 EV 充电调度模型如下：

$$\min f_1 = \frac{1}{N_e}\sum_{i=1}^{N_e}(\omega_1 T_i + \omega_2 C_i) \tag{1}$$

$$\max f_2 = \frac{1}{T_s}\sum_{t=1}^{T_s} P_{PV}(t)\Delta t \tag{2}$$

约束条件为：

$$S^{\min} \leqslant S(t) \leqslant S^{\max} \tag{3}$$

$$\frac{P_{EV}(t)}{\eta_1} + \frac{P_B(t)}{\eta_1} = P_{PV}(t)\eta_1 + P_D(t)\eta_2 \tag{4}$$

$$|P_B(t)| \leqslant P_B^{\max} \tag{5}$$

$$0 \leqslant P_D(t) < P_D^{\max} \tag{6}$$

$$\sum_{i=1}^{N_s}\varphi_i = 1 \tag{7}$$

$$\sum_{j\in W_i} u_{ij} = 1 \tag{8}$$

式中：N_e为 EV 总数量；T_i为用户 i 的总时间，包括路程时间、等待时间与充电时间，其中路程时间为用户从充电触发位置出发直至抵达目标能源站的路程耗时；C_i为用户 i 的费用，包括充电费用与服务费用；ω_1，ω_2 分别为用户时间与费用系数；T_s 为仿真总时间；$P_{EV}(t)$，$P_{PV}(t)$，$P_D(t)$，$P_B(t)$ 分别为 t 时刻站内充电负荷、光伏出力、配电网出力以及储能充放电功率，其中储能充电时 $P_B(t)$ 值为正，放电时 $P_B(t)$ 值为负；η_1，η_2 分别为站内 DC/DC 模块及 AC/DC 模块效率；$S^{\min}$，$S^{\max}$ 分别为储能系统荷电状态(state of charge，SOC)下限与上限；$P_B^{\max}$ 为储能最大充放电功率；$P_D^{\max}$ 为充电桩最大购电功率；N_s为区域范围内能源站数量；φ_i 为能源站选择变量；W_i 为所有与节点 i 相连的道路节点集合；u_{ij}为用户在节点 i 处的道路选择变量。

针对充光储能源站的 EV 调度模型属于多目标多约束优化问题，基于规划的方法以及启发式算法虽然可以进行求解，但这些算法均为离线运算且面对实际交通拓扑网络运算耗时较长。同时，不同日期下天气条件、用户充电需求等均存在较大差异，模型均需要重新求解，耗时较长且难以实现在线实时调度。

2 基于 DRL 的 EV 调度方法

2.1 DRL 基本原理

DRL 是一种结合深度学习的感知能力与强化学习的决策能力的人工智能算法。通过智能体不断与环境进行交互，并采取一定的动作使得累计奖励最大化[17—18]。智能体本质上是一个状态空间到动作空间的映射关系。强化学习算法以马尔科夫过程(Markov decision process，MDP)为数学基础，即环境下一时刻状态仅与当前状态有关，与前序状态无关。

强化学习算法采用状态-动作值函数 $Q^{\pi}(s,a)$ 来评价状态 s 时采取动作 a 的好坏，Q 函数的贝尔曼方程可表示为：

$$Q^{\pi}(s,a) = E\left(r(s,a,s') + \gamma \max_{\pi} Q^{\pi}(s',a')\right) \tag{9}$$

式中：$r(s,a,s')$为智能体采取动作 a，状态 s 转变为

s'对应的即时奖励;π 为智能体在当前状态 s 下决定下一动作 a 的策略函数;E 为数学期望;$\gamma \in [0,1]$,为折扣率,γ 接近于 0 时,智能体更在意短期回报,γ 接近于 1 时,智能体更在意长期回报。

在传统 Q 学习过程中,状态-动作-Q 值以表格的实行进行记录,智能体在状态 s 下查找 Q 表并采取最大 Q 值对应的动作 a^*。然而,实际问题中状态空间及动作空间往往很大,Q 学习方法难以实践。在 Q 学习框架基础上,深度 Q 网络(deep Q network,DQN)以深度神经网络代替 Q 表进行函数逼近[19],拟合状态-动作与 Q 值的映射关系,其贝尔曼迭代方程可表示为:

$$Q(s,a;\theta^+) = Q(s,a;\theta^+) + \alpha(r(s,a,s') + \gamma \max_{a'} Q(s',a';\theta^-) - Q(s,a;\theta^+)) \quad (10)$$

式中:$\alpha \in [0,1]$,为学习率;θ^+ 为评价网络参数;θ^- 为目标网络参数。学习过程中,评价网络每隔一定回合数将参数复制给目标网络,通过 2 个网络的配合以提高算法稳定性。

2.2 人类行为决策理论

EV 用户在充电过程中不仅仅追求预期效用的最大化,也会受限于认知水平及主观心理情绪等因素的影响,因此很难选择出全局最优或个人利益最大的充电选择方案。事实上,个体往往寻求决策后的正面情绪,从而规避决策可能带来的负面情绪。为此,文中引入后悔理论建立人类行为决策心理模型,刻画用户在 EV 充电调度过程中的心理状态,作为 DRL 智能体“人-车-站”环境状态感知的一部分。

后悔理论最早由 Bell 提出,其将后悔描述为一件给定事件的结果或状态与他将要选择的状态进行比较所产生的情绪[19]。依据人类在离散事件选择中的后悔规避心理,当所选方案优于备选方案时,决策者会感到欣喜,反之则会感到后悔。因此,决策者个体更倾向于选择预期后悔最小的方案。后悔理论通过式(11)量化决策者在选择过程中对所选方案与备选方案的感知效应[20]:

$$U_i = F_i + \sigma_i = \sum_{j=1,j\neq i}^{N_s} \sum_{k=1}^{N_a} \ln(1 + e^{\xi_k(x_{j,k} - x_{i,k})}) + \sigma_i \quad (11)$$

式中:U_i为选择方案 i 的随机效用值;F_i为选择方案 i 的可确定效用值;$x_{j,k}$ 为随机效应误差;N_s为总方案个数,即能源站个数;N_a为总属性因素个数;$x_{j,k}$ 为 j 方案在属性 k 上的取值;ξ_k 为属性 k 的估计参数,反应决策者对该属性的偏重;σ_i 为随机效用值。当 σ_i 服从独立同分布式时,决策者选择方案 i 的概率可表述为:

$$P_u(i) = \mathrm{prob}(F_i > F_j, \forall j \neq i) = \frac{e^{F_i}}{\sum_{j=1}^{N_s} e^{F_j}} \quad (12)$$

可见,后悔理论的实质是通过比较不同方案效用差 $x_{j,k} - x_{i,k}$,模拟人类在多方案选择中的思维过程,最终按照一定概率做出方案选择。文中基于后悔理论将 EV 用户参与调度总时间与总费用作为 2 个属性,将所有能源站作为方案集,通过计算用户对各方案的效用值 U_i作为智能体对环境状态感知的一部分,其具体模型如式(13)所示。

$$U_i = \sum_{j=1,j\neq i}^{N_s} (\ln(1 + e^{\xi_1(T_{sche,j} - T_{sche,i})}) + \ln(1 + e^{\xi_2(C_{sche,j} - C_{sche,i})})) + \sigma_i \quad (13)$$

式中:ξ_1,ξ_2 分别为用户对时间与费用偏重;$T_{sche,i}$ 为用户选择能源站 i 的总时间,包括路程时间、等待时间与充电时间;$C_{sche,i}$ 为用户选择能源站 i 的费用,包括充电费用与服务费用,其计算公式详见文献[21]。

2.3 DQN 实现 EV 充电调度

针对能源站的 EV 充电调度问题每一个时刻的状态仅与前一时刻状态及智能体动作有关,符合马尔科夫决策过程,因此,文中采用 DQN 方法建立 EV 充电调度模型,利用智能体进行“人-车-站”多主体状态感知,通过不断地探索与利用,建立状态-动作与 Q 值的映射关系,实现 EV 实时调度。模型中对状态、动作及奖励的定义如下。

(1) 状态。为实现智能体对环境状态的有效感知,文中定义环境状态由 EV“时-空-能量”状态、能源站“充-光-储”运行状态及用户心理状态构成,因此可建立状态 s_t,如式(14)所示。

$$s_t = (t, L_{EV,t}, E_{EV,t}, P_{EV,t}, P_{PV,t+1}, E_{B,t}, U_{U,t}) \quad (14)$$

式中:t 为当前时刻;$L_{EV,t}$ 为当前时刻 EV 位置;$E_{EV,t}$ 为当前时刻 EV 动力电池 SOC;$P_{EV,t}$ 为当前时刻各能源站 EV 的充电负荷;$P_{PV,t+1}$ 为各能源站 t+1 时刻光伏出力预测值;$E_{B,t}$ 为当前时刻各能源站储能系统 SOC;$U_{U,t}$ 为用户对各备选能源站的感知效用值。

(2) 动作。为实现 EV 的充电调度,将目标能源站与导航路径的选择作为智能体的动作,则 t 时刻智能体动作 a_t可表示为:

$$a_t = (x_{ES,t},\ x_{link,t}) \quad x_{ES,t} \in D, x_{link,t} \in L \quad (15)$$

式中:$x_{ES,t}$ 为智能体选择的能源站;$x_{link,t}$ 为智能体选择的当前道路;D 为能源站位置集合;L 为与当前道路节点相连的节点集合。

(3) 奖励。由于调度过程涉及到途中导航与到

站充电，因此可将智能体与环境交互所得的奖励分为途中奖励与到站奖励。其中，途中奖励主要考虑用户方面路程花费时间与动力电池能量代价，到站后奖励由光伏消纳功率及用户在站时间决定。

$$r_t=\begin{cases}-\lambda_1 d_{ij}\alpha-\lambda_2 d_{ij}/v_{ij} & L_{\mathrm{EV},t}\neq x_{\mathrm{ES},t}\\ \delta_1\bar{P}_{\mathrm{PV},t}-\delta_2(T_{\mathrm{wait}}+T_{\mathrm{charge}}) & L_{\mathrm{EV},t}=x_{\mathrm{ES},t}\end{cases} \tag{16}$$

式中：$L_{\mathrm{EV},t}$ 为当前时刻 EV 位置；$x_{\mathrm{ES},t}$ 为目标能源站位置；d_{ij} 为道路节点 i 至 j 的距离；v_{ij} 为道路节点 i 至 j 的平均行驶速度；α 为 EV 单位距离耗电量；λ_1，λ_2 分别为能耗与时间奖励系数；T_{wait}，T_{charge} 分别为 EV 在站等待时间与充电时间；$\bar{P}_{\mathrm{PV},t}$ 为各能源站平均光伏消纳功率；δ_1，δ_2 分别为能源站光伏消纳系数与用户充电时间代价系数。

由于智能体在学习前期缺少历史样本，如果采用确定性的贪心策略进行动作选择，容易造成局部收敛甚至不收敛。因此，文中引入时变 ε-greedy 策略，在前期的学习中增大智能体探索能力，在后期的学习中有效利用前期历史样本进行决策，如式(17)所示。

$$a_t=\begin{cases}\mathrm{random}\ A & \beta<(N-n)\varepsilon/N\\ \arg\max\limits_{a_t\in A}Q(s_t,a_t) & \beta\geqslant(N-n)\varepsilon/N\end{cases} \tag{17}$$

式中：N 为总训练回合数；n 为当前训练回合数；β 为(0,1)随机数；ε 为比例参数；random 为随机函数，即从 A 中随机抽取动作；arg max 为求解函数值最大化，即返回使得 Q 值最大的动作。因此，在训练前期智能体有更大概率是从动作集合 A 中随机选取动作，而在训练中后期，则更有可能选取最优动作。同时，时变 ε-greedy 策略逐步减小 ε，可以提高算法的收敛速度。

3 EV 充电调度框架

基于 DRL 的 EV 充电调度实现框架如图 2 所示。该过程可分为以下 3 个步骤：

(1) 智能体通过更新时间、EV 位置及动力电池 SOC 获取车辆状态，更新各能源站运行状态并预测下一时刻光伏出力，通过后悔理论感知 EV 用户的心理状态，得到当前时刻环境状态 s_t。

(2) 智能体将感知到的环境状态输入深度神经网络，得到各备选动作的 Q 值，通过时变 ε-greedy 策略选择动作 a_t。

(3) 智能体执行所选动作，重复上述步骤直至车辆抵达所选能源站。

4 算例分析

4.1 参数设置

为验证文中所提策略的有效性与实用性，选取南京市部分区域，范围为经度(东经)118.735 152~118.784 076，纬度(北纬)32.059 057~32.092 003 作为算例路网。同时，选取该区域已经投入运营的 15 座能源站，假设该区域能源站均配置了光伏发电及储能系统，且站内充电桩均为快充，具体配置详见表 1。

根据文献[22]EV 出行规律，文中在该区域一天中引入 1 000 辆 EV，设 EV 动力电池容量为 40 kW·h，并设初始 SOC 服从对数均值为 3.2，对数标准差为 0.48 的对数正态分布。考虑电池充放电深度对其寿命的影响，取 EV 结束充电时的终止 SOC 均为 90%。

4.2 智能体训练过程

设置 DQN 算法中智能体学习率 $\alpha=0.85$，奖励折扣率 $\gamma=0.85$，ε-greedy 策略中 ε 初值为 0.5，每回

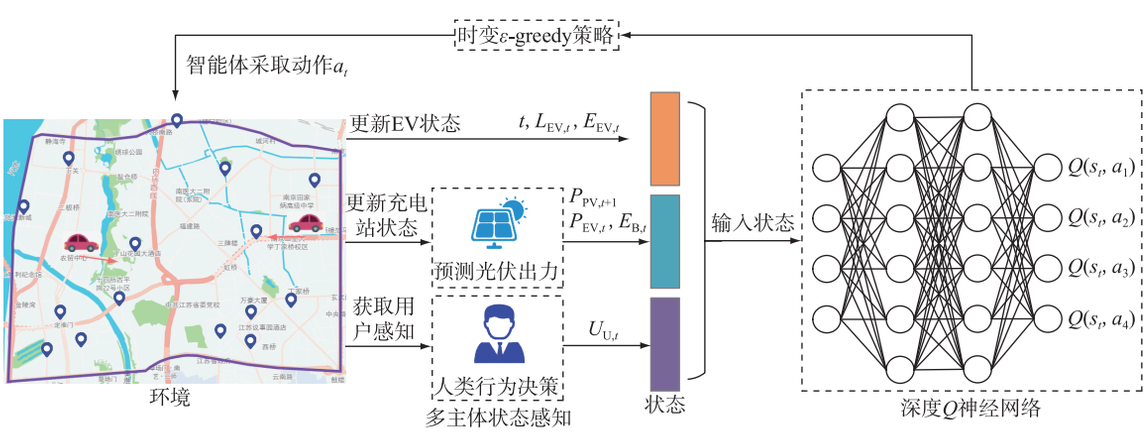


图 2 优化调度策略实现流程

Fig.2 Flow chart of optimized scheduling strategy

表 1 能源站基本参数表

Table 1 Basic parameters of energy station

系统参数	数值
光伏设备功率/kW	36
光伏设备效率	0.9
储能设备容量/(kW·h)	300
储能充放电功率/kW	120
储能设备效率	0.9
储能 SOC 下限	0.1
储能 SOC 上限	0.85
充电桩数量/个	10
充电桩功率/kW	60
充电桩效率	0.9

合递减 7.5×10^{-4} 直至为 0,Q 网络采用 150×120 全连接神经网络。总训练回合数设置为 4 000 次,可得训练过程中智能体训练过程中平均奖励值如图 3 所示。

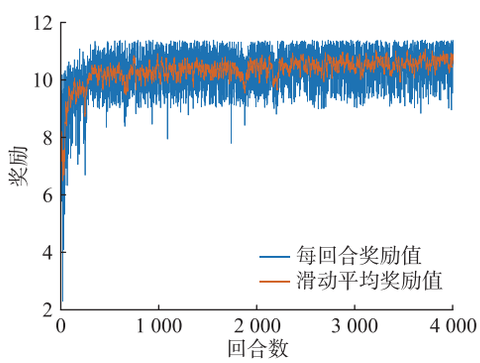


图 3 训练过程奖励值

Fig.3 Reward value of training process

由图 3 可知,在训练前期智能体每回合奖励呈现一个明显的上升阶段,并在 500 回合左右实现收敛,奖励值稳定于 10.44。这是因为 ε-greedy 策略的存在,使得智能体在前期能够不断探索环境,而当 $n=500$ 时,$(N-n)\ \varepsilon/N=0.11$,表明 500 回合之后智能体更大概率是根据当前学习到的历史经验进行动作选择。由于每一回合中 EV 初始时空分布存在差异,且光伏出力存在一定波动,所以智能体所得奖励存在一定波动,但训练后期平均奖励明显高于训练前期,表明智能体已拟合状态-动作与 Q 值的映射关系,并能够进行最优动作的选取。

4.3 泛化能力分析

为分析所提 DRL 算法泛化能力,考虑能源站日常运行状态,设置晴天、突变天气及阴雨天光伏出力如图 4 所示,其中红色宽带为光伏出力概率区间。设置训练 1~1 000 回合对应晴天,1 001~2 000 回合对应突变天气,2 001~3 000 回合对应阴雨天气,可得训练奖励如图 4 所示。

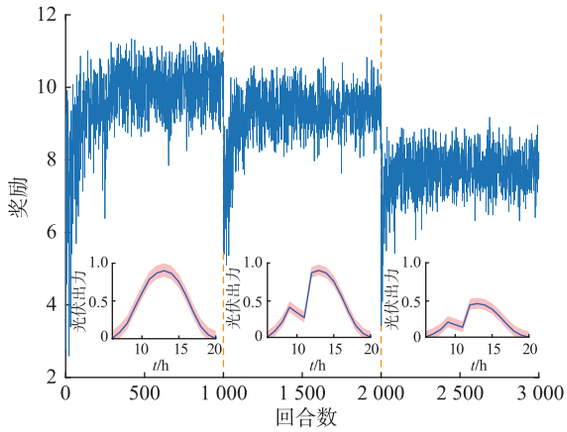


图 4 考虑泛化能力的训练奖励

Fig.4 Training reward considering generalization ability

由图 4 可知,不同天气类型对智能体所获得奖励值有较大影响,3 种天气下智能体平均奖励分别为 9.95,9.38,7.23,特别地,阴雨天气奖励值较晴天降低 27.34%。这是由于智能体的到站奖励与区域内能源站平均光伏消纳功率有较大关系,虽然阴雨天气智能体所得奖励较晴天更低,但此时智能体已经实现了最优策略的学习。同时,观察算法收敛速度可见,所提 DQN 方法在前 2 种场景下分别在 400 与 200 回合达到稳定,而在第 3 种场景下训练约 80 回合即实现收敛,表明智能体能够有效利用前期累积的经验,当环境状态发生较大改变时,其能够调整神经网络参数以快速适应当前环境状态。

进一步,在上述 3 种场景下,EV 分别采取无序充电及文中所提 DQN 方法所得光伏消纳率如表 2 所示。

表 2 不同场景光伏消纳率

Table 2 Objective value of different scheduling scale %

光伏消纳率	无序充电	DQN 方法
场景 1	75.34	81.33
场景 2	77.83	83.90
场景 3	86.32	98.05
平均值	79.82	87.76

从表 2 可见,在场景 1 中,无序充电情况下各能源站平均光伏消纳率仅为 75.31%,而文中 DQN 方法只涉及 EV 用户对能源站的选择及导航问题,在时间维度不存在调度关系,因此基于 DQN 方法的光伏消纳率也仅提高了 6.02%。3 种不同场景下文中所提方法平均提高光伏消纳率7.94%,其中场景 3 效果最为明显,提高 11.73%。可见,所提方法能够适应不同场景下的能源站运行状态,有效提高光伏消纳水平。

4.4 算法实时性分析

进一步地，为了分析所提 DQN 方法的计算效率以及实时性，文中将常规的规划方法和启发式算法与 DQN 算法进行比较。文中所提 EV 调度问题可以采用商业 Cplex 求解器以及粒子群优化算法（particle swarm optimization，PSO）进行求解。为体现算法在实际应用中是实时性，不同求解方法的单辆 EV 平均计算耗时如表 3 所示。

表 3 不同算法计算耗时对比

Table 3 Comparison of computation time of different algorithms s

算法	耗时
Cplex	11.27
PSO	22.35
DQN	0.007 1

由表 3 可知，训练好的 DQN 模型在计算速度上具有较大优势。PSO 通过粒子群逐步迭代寻优，计算结果可能收敛于局部最优。同时，每次求解重复迭代直至收敛的过程，使得 PSO 的决策时间较长。当环境状态发生改变时，传统的优化算法均需要重新进行优化求解，而 DQN 模型只需将当前时刻的环境状态作为输入，通过训练好的网络即可得到 EV 的动作输出，能够在毫秒级完成调度策略的制定，满足实时调度的需求。

4.5 非理性人心理分析

上述智能体训练过程中，后悔理论中 EV 用户对时间成本与费用成本的感知系数均为 0.5。为探究人类非理性状态感知对智能体决策的影响，分别定义 2 种非理性人：非理性人 1 更在意费用成本（$\xi_1 = 0.2, \xi_2 = 0.8$）；非理性人 2 更在意时间成本（$\xi_1 = 0.8, \xi_2 = 0.2$），分别与最短路径法导航结果相比较，图 5 给出了不同非理性人在同一起讫点时模型所推荐的导航路径。

由图 5 可知，针对 2 种非理性人，智能体共选取出 7 条路线，其中均包含了最短路径。对于非理性人 1，智能体共推荐出行驶路线 5 条，平均路程 4.37 km，平均行驶时间 8.54 min。对于非理性人 2，智能体共推荐路线 7 条，平均路程 4.62 km，较前者增长 5.72%，平均行驶时间 8.61 min，较前者增加 0.82%。通过对比可知，若用户表现出更在意时间成本，智能体则会更倾向于具有探索精神，以极小的时间代价，进而探索可能的最佳路线。可见，由于不同行为人在后悔理论中对各因素感知权重不同，智能体能够通过状态感知获取 s_t，并在训练过程中不断学习与调整 Q 网络参数与映射关系，实现考虑用户异质性的 EV 充电导航与调度。

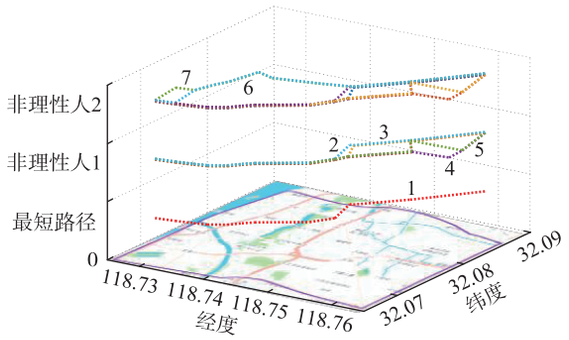
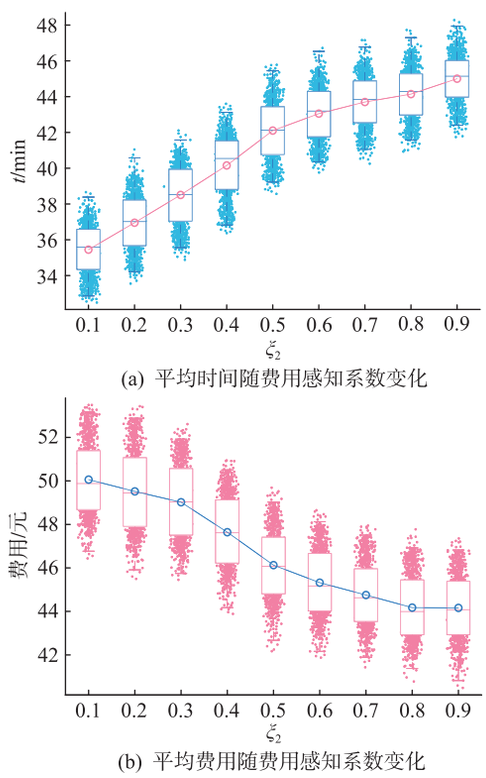


图 5 不同情况下导航路径

Fig.5 Navigation path in different situations

最后，为探究不同非理性人心理状态对智能体调度策略的影响，分别设用户的费用感知偏重 $\xi_2 = 0.1, 0.2, \cdots, 0.9$（时间感知偏重 $\xi_1 = 0.9, 0.8, \cdots, 0.1$），可得基于 DRL 方法的用户平均时间与费用变化曲线如图 6 所示。



(a) 平均时间随费用感知系数变化

(b) 平均费用随费用感知系数变化

图 6 不同感知偏重对用户影响

Fig.6 The impact of different perception bias on users

由图 6 可知，随着用户费用感知系数的增大，用户平均费用逐渐减小，平均用时逐渐增大。特别地，当 $\xi_2 = 0.1$，即用户特别在意时间成本时，此时平均耗时 35.44 min，平均费用 50.06 元；当 $\xi_2 = 0.9$，即用户特别在意费用时，平均耗时 45.01 min，较前者增加了 27%，而平均费用 44.16 元，较前者降低了 11.79%。由时间与费用变化趋势可以看出，不同特

质车主对于充电所用时间与费用的预期存在一定差异,当费用感知系数每增加 0.1 时,用户费用平均降低 1.55%,而时间感知系数每增加 0.1 时,用户时间平均降低 2.93%。可见,EV 用户对于充电过程所用时间感知更为敏感。

5 结论

针对能源站 EV 充电导航与调度问题,提出基于 DRL 方法的调度策略。算例从多角度分析了优化调度策略,得到如下结论:(1) DQN 方法中智能体对 EV 状态、能源站运行状态以及用户心理状态进行全状态感知,通过学习状态-动作与 Q 值的映射关系能够有效进行充电调度。(2) 在晴天与阴雨天等能源站常见运行场景下,所提方法均能够兼顾用户心理感知进行调度,同时有效提高了能源站光伏利用率,具有较强的实用性与泛化能力。(3) 不同行为人对时间与费用的感知效用会影响智能体状态感知与策略参数,进而影响所提方法对其的导航与调度策略。

尽管如此,限于篇幅文中并未分析 DQN 算法参数对调度策略的影响,在下一步的工作中 DQN 算法参数的选择可以继续完善。此外,基于用户感知异质性的研究,可以进一步改进所提策略。

参考文献:

[1] 肖定垚,王承民,曾平良,等. 电力系统灵活性及其评价综述[J]. 电网技术,2014,38(6):1569-1576.
XIAO Dingyao, WANG Chengmin, ZENG Pingliang, et al. A survey on power system flexibility and its evaluations[J]. Power System Technology,2014,38(6):1569-1576.

[2] 刘洪,阎峻,葛少云,等. 考虑多车交互影响的电动汽车与快充站动态响应[J]. 中国电机工程学报,2020,40(20):6455-6468.
LIU Hong,YAN Jun,GE Shaoyun,et al. Dynamic response of electric vehicle and fast charging stations considering multi-vehicle interaction[J]. Proceedings of the CSEE,2020,40(20):6455-6468.

[3] 邵尹池,穆云飞,余晓丹,等. "车-路-网"模式下电动汽车充电负荷时空预测及其对配电网潮流的影响[J]. 中国电机工程学报,2017,37(18):5207-5219,5519.
SHAO Yinchi,MU Yunfei,YU Xiaodan,et al. A spatial-temporal charging load forecast and impact analysis method for distribution network using EVs-traffic-distribution model[J]. Proceedings of the CSEE,2017,37(18):5207-5219,5519.

[4] 江明,许庆强,季振亚. 基于时序差分学习的充电站有序充电方法[J]. 电力工程技术,2021,40(1):181-187.
JIANG Ming, XU Qingqiang, JI Zhenya. Coordinated charging approach for charging stations based on temporal difference learning[J]. Electric Power Engineering Technology,2021,40(1):181-187.

[5] KORKAS C D,BALDI S,YUAN S,et al. An adaptive learning-based approach for nearly optimal dynamic charging of electric vehicle fleets[J]. IEEE Transactions on Intelligent Transportation Systems,2018,19(7):2066-2075.

[6] CAI H,CHEN Q Y,GUAN Z J,et al. Day-ahead optimal charging/discharging scheduling for electric vehicles in microgrids[J]. Protection and Control of Modern Power Systems,2018,3(1):1-15.

[7] 阎怀东,马汝祥,柳志航,等. 计及需求响应的电动汽车充电站多时间尺度随机优化调度[J]. 电力系统保护与控制,2020,48(10):71-80.
YAN Huaidong, MA Ruxiang, LIU Zhihang, et al. Multi-time scale stochastic optimal dispatch of electric vehicle charging station considering demand response[J]. Power System Protection and Control,2020,48(10):71-80.

[8] 高昇宇,柳志航,卫志农,等. 城市智能光储充电塔自适应鲁棒日前优化调度[J]. 电力系统自动化,2019,43(20):39-48.
GAO Shengyu, LIU Zhihang, WEI Zhinong, et al. Adaptive robust day-ahead optimal dispatch for urban smart photovoltaic storage and charging tower[J]. Automation of Electric Power Systems,2019,43(20):39-48.

[9] 李睿雪,胡泽春. 电动公交车光储充电站日运行随机优化策略[J]. 电网技术,2017,41(12):3772-3780.
LI Ruixue,HU Zechun. Stochastic optimization strategy for daily operation of electric bus charging station with PV and energy storage[J]. Power System Technology, 2017, 41(12): 3772-3780.

[10] 肖浩,裴玮,孔力. 含大规模电动汽车接入的主动配电网多目标优化调度方法[J]. 电工技术学报,2017,32(S2):179-189.
XIAO Hao, PEI Wei, KONG Li. Multi-objective optimization scheduling method for active distribution network with large scale electric vehicles[J]. Transactions of China Electrotechnical Society,2017,32(S2):179-189.

[11] 杜明秋,李妍,王标,等. 电动汽车充电控制的深度增强学习优化方法[J]. 中国电机工程学报,2019,39(14):4042-4049.
DU Mingqiu, LI Yan, WANG Biao, et al. Deep reinforcement learning optimization method for charging control of electric vehicles[J]. Proceedings of the CSEE,2019,39(14):4042-4049.

[12] ZHANG C,LIU Y,WU F,et al. Effective charging planning based on deep reinforcement learning for electric vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2021,22(1):542-554.

[13] 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 dispatch strategy for electric vehicles based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2020, 44(22): 161-167.

[14] WANZ Q, LI H P, HE H B, 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.

[15] QIAN T, SHAO C C, WANG X L, et al. Deep reinforcement learning for EV charging navigation by coordinating smart grid and intelligent transportation system[J]. IEEE Transactions on Smart Grid, 2020, 11(2): 1714-1723.

[16] 何阳,张宇,王育飞,等. 考虑负荷优化的电动汽车光伏充电站储能容量配置[J]. 现代电力,2019,36(5):76-81.
HE Yang, ZHANG Yu, WANG Yufei, et al. Energy storage capacity configuration of PV-integrated EV charging station considering load optimization[J]. Modern Electric Power, 2019, 36(5): 76-81.

[17] 张自东,邱才明,张东霞,等. 基于深度强化学习的微电网复合储能协调控制方法[J]. 电网技术,2019,43(6):1914-1921.
ZHANG Zidong, QIU Caiming, ZHANG Dongxia, et al. A coordinated control method for hybrid energy storage system in microgrid based on deep reinforcement learning[J]. Power System Technology, 2019, 43(6): 1914-1921.

[18] 刘全,翟建伟,章宗长,等. 深度强化学习综述[J]. 计算机学报,2018,41(1):1-27.
LIU Quan, ZHAI Jianwei, ZHANG Zongchang et al. A survey on deep reinforcement learning[J]. Chinese Journal of Computers, 2018, 41(1): 1-27.

[19] BELL D E. Regret in decision making under uncertainty[J]. Operations Research, 1982, 30(5): 961-981.

[20] 高玉芳. 基于后悔理论的城市轨道交通动态配流模型研究[D]. 北京:北京交通大学,2018.
GAO Yufang. Research on dynamic assignment model of urban rail transit based on regret theory[D]. Beijing: Beijing Jiaotong University, 2018.

[21] 邢强,杨祺铭,范军太,等. 基于数据驱动方式和行为决策的电动汽车快充需求预测模型[J]. 电网技术,2020,44(7):2439-2453.
XING Qiang, YANG Qiming, FAN Juntai, et al. Electric vehicle fast charging demand forecasting model based on data-driven approach and human behavior decision-making[J]. Power System Technology, 2020, 44(7): 2439-2453.

[22] 程骏. 电动汽车充电站运行调度策略研究[D]. 南京:东南大学,2016.
CHENG Jun. Research on operation scheduling strategy for electric vehicle charging station[D]. Nanjing: Southeast University, 2016.

作者简介:


孙广明

孙广明(1979),男,硕士,高级工程师,从事电动汽车充换电设施监控与运营管理工作(E-mail:alex092416@163.com);

陈良亮(1989),男,博士,研究员级高级工程师,从事电动汽车充换电技术相关工作;

王瑞升(1998),男,硕士在读,研究方向为电动汽车与电网互动技术研究。

A deep reinforcement learning-based scheduling strategy of photovoltaic-storage-charging integrated energy stations

SUN Guangming[1], CHEN Liangliang[1], WANG Ruisheng[2], CHEN Zhong[2], XING Qiang[2]

(1. NARI Group (State Grid Electric Power Research Institute) Co., Ltd., Nanjing 211106, China;
2. School of Electrical Engineering, Southeast University, Nanjing 210096, China)

Abstract: Large-scale electric vehicles (EVs) scheduling models are complex and require high calculation capacity. To solve these problems, machine learning methods have attracted more and more attention in electric vehicle charging and navigation scheduling. For the photovoltaic-storage-charging integrated energy station, a scheduling strategy of the energy stations based on deep reinforcement learning (DRL) is proposed in this paper. Firstly, the operation strategy of energy station and the basic theory of deep reinforcement learning are analyzed. Secondly, the users psychological state of time and cost for different charging schemes are described based on regret theory, and the agent perception model of user-EV-station state environment is established. To improve the convergence speed of the algorithm, time varying ε-greedy strategy is introduced as action selection method of agent. Finally, multi-scenario simulations are designed based on the actual road network and energy stations in Nanjing. The results show that the proposed method effectively improves the photovoltaic consumption rate of the energy station under the condition of considering the psychological effect of various users. The proposed method provides a new idea for electric vehicle charging scheduling.

Keywords: electric vehicle; photovoltaic-storage-charging integrated energy station; charging scheduling; deep reinforcement learning; regret theory; full perception model

(编辑　李栋)