基于 A3C 的有序充电算法

张文龙,张 洁
(南京邮电大学 计算机学院,江苏 南京 210023)

摘 要:由于电动汽车的日益普及,其充电问题已成为电力系统的新的用电挑战。实际生活中,充电站一般都被认为是电动汽车有序充电行为的调度主体。为解决传统模型驱动的充电算法无法应用于电动汽车随机进站的问题,提出将数据驱动的无模型深度强化学习算法 A3C(Asynchronous Advantage Actor-critic,异步演员评论家算法)应用于有序充电。该算法利用特征函数来近似模型所需要的价值函数和策略函数,解决因随机进站而引起的空间维度变化的问题。通过需求响应机制关联充电费用和需求,实现两者的动态调度。为避免因为经验回放而导致的数据相关性过强,利用多线程实现模型与多个环境进行互动,提高了模型的收敛性。最后以某地区充电站实测数据为例进行仿真分析。结果表明,该算法在只基于历史充电数据的情况下能优化充电行为,较大程度地抑制充电负荷方差,实现削峰填谷,同时在满足用户需求的基础上提高充电站收益。

关键词:有序充电;数据驱动;强化学习;深度学习;A3C
中图分类号:TP391.9 **文献标识码**:A **文章编号**:1673-629X(2023)01-0173-05
doi:10.3969/j.issn.1673-629X.2023.01.026

Orderly Charging Algorithm Based on A3C

ZHANG Wen-long,ZHANG Jie
(School of Computer Science,Nanjing University of Posts and Telecommunications,Nanjing 210023,China)

Abstract:Due to the increasing popularity of electric vehicles (EV),the charging problem has been a new challenge of electrical system. In particular,charging stations are always considered as an important role who schedule the orderly charging behavior of EV. In order to solve the problem that conventional model-driven charging algorithms cannot be applied to the situation where electric EV enter the station randomly,propose to apply a data-driven model-free reinforcement learning algorithms A3C (Asynchronous Advantage Actor-critic) for orderly charging. The algorithm deal with the varying state spaces caused by random EV arrivals by approximating the state function and policy function with feature function. The demand response mechanism is applied to associate the charging price with the charging demands and dynamic scheduling them. To avoid the strong correlation caused by experience replay,multiprocessing is used to implement the effect that the model interact with multiple environments through which can improve the convergence of the algorithm. Finally,the simulation analysis is conducted by the measured data of charging stations in a certain area. The results show that the purposed algorithm can optimize the charging behavior even the only known is the previous charging data,reduce the charging load variance greatly and realize the peak load shifting of the grid. Beside satisfy the EVs demand,it can also increase charging station's profits.

Key words:orderly charging;data-driven;reinforcement learning;deep learning;A3C

0 引 言

随着电动汽车渐渐成为人们主流的出行选择,其充电调度的重要性也渐渐凸显,大量充电汽车接入电网势必会造成电荷显著改变,给电网带来谷峰差距过大、变压器过载等潜在威胁,因此,对充电负荷进行有序调控是十分必要的[1]。充电站因其聚合控制的优势通常是实施充电调度的主体,对充电站汽车有序充电的研究一直是智能用电领域的重要议题[2-3]。传统的有序充电调度算法是模型驱动的,例如,文献[4-6]提出了基于粒子群优化、思维进化、拉格朗日松弛优化等算法的充电调度策略,但是模型驱动算法不能反映充电行为的不确定性,算法的好坏完全由先验模型假设的好坏决定,在实际应用中如果要确定好的先验模型假设需要经过大量的采样统计,成本过高,应用范围受

收稿日期:2022-02-14 **修回日期**:2022-06-15
基金项目:国家重点研发计划(2018YFB1500902);南京邮电大学校级科研基金(NY219122)
作者简介:张文龙(1996-),男,硕士研究生,研究方向为电动汽车有序充电和电动汽车充电站运行;通信作者:张 洁(1981-),女,高级工程师,硕导,博士,研究方向为电动汽车与电网互动、电力信息集成。

限。使用数据驱动的算法是有序充电策略的另一思路,其最大的优势是不需要先验模型和假设,完全依赖于对过去和当前的数据观察,随着算力发展和原始充电数据的不断积累已经广泛应用于有序充电。其中无模型强化学习框架日益引起人们的重视,文献[7-8]实现了基于 Q-learning 和 DQN(Deep Q Network,深度 Q-learning)算法的充电调度,通过调控充电站总功率实现了电动汽车的充电控制优化;文献[9]提出一种基于 MC(Monte Carlo,蒙特卡洛)算法的多电站充电调度策略并且对电站数量有很好的扩展性;文献[10]以 MFRL(Model Free Reinforcement Learning,无模型强化学习)框架适应充电负荷的不确定性,保证所有充电行为都能及时得到满足,避免受到惩罚。除了充电调度之外,对设计充电响应需求机制的研究也在不断完善,响应机制即电动汽车会根据充电站的收费调整充电需求,文献[11-13]分别面向充电站和汽车集群设计了充电响应机制,研究表明合理的价格制定和充电调度有利于用户充电和电网系统。但是上述基于无模型强化学习的有序充电调度都没有考虑到汽车进站的随机性造成的影响,都侧重于提前一天计划好未来车辆的进出站,也没有考虑不同的充电费用对充电需求的影响,没有考虑同时调控充电费用和充电功率。

在文献[14]的基础上,该文以减少充电负荷方差和提高充电站收益为训练目标,提出基于深度强化学习 A3C 的充电站有序充电调度算法。文献[14]提出的 Sarsa 算法需要一张额外表格存储维护之前的数据纪录来进行模型的更新,而 A3C 算法结合了深度神经网络,通过对网络梯度的更新就能求出最优策略;为了解决由汽车进站的随机性引起的输入维度变化,使用 5 个特征函数来近似价值函数和策略函数,有效减少输入维度。借助充电需求响应机制关联调控充电费用和功率,实现两者的动态实时调度。之后使用某地区充电站一个月的历史充电纪录进行仿真分析,验证该算法能有效抑制充电负荷方差,在满足用户需求的基础上提高充电站收益和削峰填谷。

1 模型建立

1.1 充电站模型

该文所假设的整个充电系统的运行情况如图 1 所示。

在一个被划分成 T 个时间段的时间区间,每个时间段都会有随机车辆进站定义为 $I_t, t = 1,2,\cdots,T$,a_i、p_i、d_i 分别代表车辆 i 的进站时间、最大停留时间和充电需求。I_t 内的车辆会根据当前的充电费用确定充电需求,将充电请求交给电站,然后加入充电队列等待充电站的调度。J_t 表示在时间段 t 之前已进入充电站且还没有完成充电的车辆,在每个时间段开始会从 J_t 中选择优先级高的车辆进行充电,充电的功率还受到以下限制。

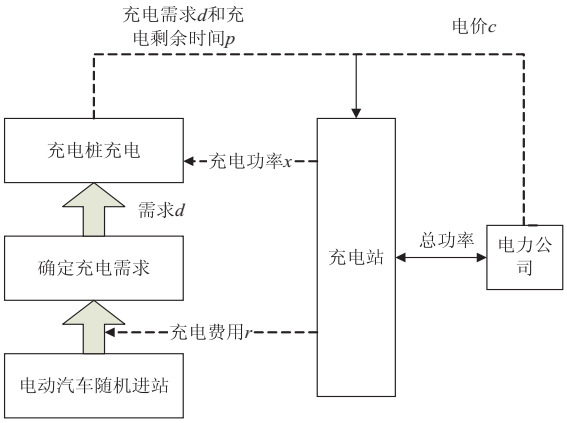

图 1 充电站模型

$$x_{i,t} \leqslant x_{i,\max}, \forall i \tag{1}$$

$$\sum_{i \in J_t} x_{i,t} \leqslant e_{\max}, t = 1,2,\cdots,T, i \in J_t \tag{2}$$

$$\sum_{t=a_i}^{a_i+p_i} x_{i,t} \geqslant d_i, \forall i \tag{3}$$

其中,$x_{i,\max}$ 和 $e_{\max}$ 分别代表车辆的最大充电功率和充电站最大充电功率,所有车辆的需求应该在离开充电站前得到满足。在充电站的整个工作过程中,充电费用和充电功率都是实时变化调整的,但对具体的车辆而言,在 t 时段进站的车辆对应的充电费用为 r_t,直到出站前都不再变化。车辆 i 的充电需求与进站时的充电费用满足 $d_i = D_i(r_t)$,D_i 是车辆 i 对应的需求响应机制函数,只与车的种类有关。由于电动汽车到达时间和充电费用的不确定性,充电站只知道已经到达的电动汽车的充电情况以及目前为止的电价变化。充电站收益在每个时段都受两部分直接影响,一是充电站向各时段新进站电动汽车收取的费用,二是充电站各时段根据充电需求向电力公司购买电力的费用,两者的差额就是充电站每个时段的收益。

1.2 充电过程模型

在每个时段开始,充电站都会基于对历史充电行为的观察包括当前充电费用和充电请求,选择最优充电功率和充电费用,这些决策的制定又会影响到未来决策的制定,因此用 MDP(Markov Decision Process,马尔可夫决策过程)来模拟整个充电过程求解最优决策[15]。目前 MDP 在有序充电领域正获得广泛应用,MDP 由状态 S、动作 A、转移函数 P 和奖励函数 R 共同组成。若上述四要素均已知,就可以通过动态规划算法求解,但实际应用中转移函数往往难以直接求出具体模型,因此一般的模型驱动算法在面临复杂问题或者模型未知时,都是选择基于对转移函数或未知模

型的先验假设来进行求解的,最终结果的好坏直接取决于先验假设建模,容易面临性能瓶颈。该文所采用的是另一种无模型数据驱动的思路,不依赖先验模型假设,只基于对历史和当前数据的观察。下面介绍由 1.1 节充电模型抽象成 MDP 后的四要素。

1.2.1　*状态 S*

t 时段的状态表示为 $S_t = (d_{i,t}, p_{i,t}, i \in J_t)$,其中 $d_{i,t}$ 表示汽车 i 在 t 时段的充电需求,$p_{i,t}$ 表示车辆 i 的最大剩余充电时间,若 $p_{i,t} \leqslant 0$, $d_{i,t} \geqslant 0$ 就表明这辆车的充电需求没有被满足,$p_{i,t} > 0$, $d_{i,t} \leqslant 0$ 则表明车辆 i 的充电需求已被满足可以从充电队列中删除。在每个时间段开始都会根据当前的充电情况更新状态 S,除了把新进站的汽车加入,对充电需求无法按时得到满足或者提早得到满足的汽车都应做相应的处理更新。

1.2.2　*动作 A*

$A_t = (r_t, x_{i,t}, \forall i \in I_t \cup J_t)$,$r_t$ 是 t 时段的充电费用,$x_{i,t}$ 是车辆 i 在 t 时段的充电功率。车辆 i 在执行对应的动作后会发生如下变化:

$$d_i = d_i - x_{i,t}, i \in J_t \tag{4}$$

$$p_i = p_i - 1 \tag{5}$$

动作维度和状态维度一样,受站内电动汽车数量影响,不仅空间维度过高,还会随充电站内车辆的数量的变化而变化,无法直接输入模型使用。根据文献[14]的证明,高维的动作空间 $A_t = (r_t, x_{i,t}, \forall i \in I_t \cup J_t)$ 可以简化成 $A_t = (r_t, e_t, \forall i \in I_t \cup J_t)$,$e_t$ 是 t 时段的总充电功率 $e_t = \sum x_{i,t}$,这样动作函数维度就固定成 2 维,同时维度的减少不会影响决策本身的可行性。

1.2.3　*转移函数 P*

在每个时间段开始,充电站会从 J_t 中选择优先级高的车辆进行充电,那么,J_{t+1} 就是由经历充电 J_t 后仍未完成充电且目前仍可能完成充电车辆和新到达车辆 I_t 共同组成。下一个状态可以表示为 $S_{t+1} = P(S_t, A_t, I_t, c_t) = (d_{i,t+1}, p_{i,t+1}, i \in J_{t+1})$,状态的变化直接可以通过观察得到,不必通过转移函数求解。

1.2.4　*奖励函数 R*

使用直接观测到的利润函数作为奖励函数。通过 1.1 节分析,每个时段的利润具体表示为 $R(S_t, A_t) = \sum r_t D_i(r_i) - c_t e_t$。该文提出的模型的优化目标之一就是使充电站利润最大化。

2　A3C 算法和特征函数

2.1　A3C

A3C 算法本质是 AC(Actor-Critic,演员-评论家)算法,Actor 部分负责生成动作并和环境交互,而 Critic 部分负责评估 Actor 的表现,并指导 Actor 下一阶段的动作。Actor 部分利用的是一类基于策略(policy based)的强化学习算法;Critic 部分使用的是一类基于价值(value based)的强化学习算法,Actor-Critic 算法结合了两种强化学习算法的优点。Actor 和 Critic 两部分的更新都与 TD(temporal-difference,时序差分)误差相关,Actor 为避免正数陷阱,通过 TD 误差更新策略;Critic 部分的损失就是 TD 误差,它的更新目标就是最小化 TD 误差[16]。

A3C 最大的不同就是放入多线程中同步训练,具体工作流程如图 2 所示。A3C 主要架构是由 Global Network(全局网络)和 worker(工人)组成的,全局网络和 worker 拥有一样的 Actor-Critic 网络结构,目的是模拟多个智能体与环境互动,全局网络并不直接参加和环境的互动,而是把自己当前学习到的最新参数共享给多个 worker,让 worker 与不同的环境进行交互,最后各 worker 将自己探索学习到的梯度汇报给全局网络进行更新[16]。

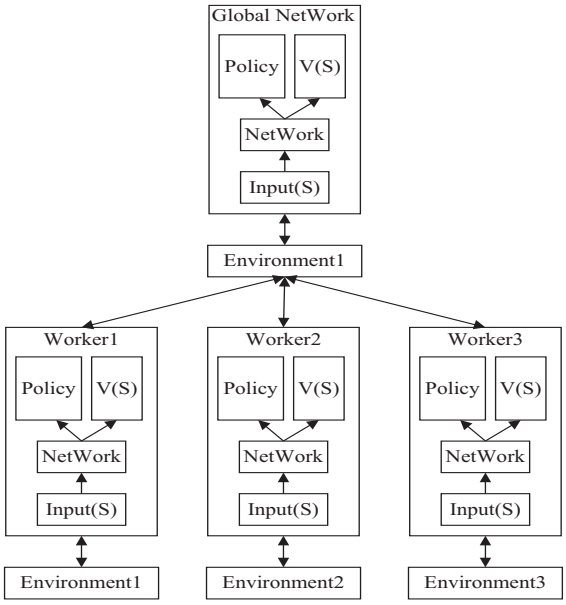

图 2　A3C 模型

A3C 完整算法:

输入:公共部分的 A3C 神经网络结构,其对应参数为 θ, ω 分别属于 Actor 和 Critic 部分;本地线程的 A3C 神经网络结构,对应的参数为 θ', ω' 全局共享的迭代次数 T,全局最大迭代次数为 T_{max};线程内单次迭代时间序列最大长度 T_{local},状态特征维度 n,动作集 Action,步长 α, β;衰减因子 γ。

a)更新时间序列 $t = 1$

b)重置 Actor 和 Critic 部分的梯度更新;将公共 A3C 的参数同步到本地 A3C 神经网络 $\theta' \leftarrow \theta$, $\omega' \leftarrow \omega$。

c)初始化 S_t, $t_{start} = t$。

d)基于策略近似函数 $\pi(a_t \mid S_t, \theta')$ 选择动作 a_t 并执行,得到回报 R_t 和下一个状态 S_{t+1}, $t = t + 1$, $T = T + 1$。

e)如果 S_t 是终止状态或者 $t - t_{start} = T_{local}$,执行步骤 f,否则,返回步骤 d。

f) 计算 S_t 的价值函数 $V = \begin{cases} 0, s_t \text{ 为终止状态} \\ v(s_t, \omega'), \text{else} \end{cases}$

g) for $i \in (t-1, t-2, \cdots, t_{start})$:

$V = R_i + \gamma V$

累计 Actor 的本地梯度更新:

$d\theta \leftarrow d\theta + \nabla_\theta \log \pi(a_i \mid s_i, \theta')(V - v(s_i, \omega'))$

累计 Critic 的本地梯度更新:

$$d\omega \leftarrow d\omega + \frac{\partial (V - v(s_i, \omega'))^2}{\partial \omega'}$$

h) 更新全局神经网络的模型参数: $\theta = \theta - \alpha d\theta$, $\omega = \omega - \beta d\omega$ 。

i) 如果 $T \geqslant T_{max}$,退出算法,否则,返回步骤 b。

2.2 特征函数

2.1 节提到 Actor 部分和 Critic 部分通常都是使用神经网络建立模型,两者的输入都是 MDP 中的状态 S ,但是汽车进站的随机性导致状态 S 维度过高且不固定,无法直接作为神经网络的输入。为了使输入维度固定,使用状态的特征函数作为神经网络输入,特征函数主要基于目标函数和约束条件来构造,具体使用下面 5 个特征函数[14]。

$$f_1(s_t) = \sum_{t=1}^{T} r_t D_i(r_t), i \in I_t \cup J_t \tag{6}$$

$$f_2(s_t) = -c_t e_t \tag{7}$$

$$f_3(s_t) = -\sum_{t=1}^{T} (P_t - p_{i,t}) \theta_1 \sum d_{i,t}, i \in I_t \cup J_t \tag{8}$$

$$f_4(s_t) = -\sum \theta_2^{p_{i,t}} \sum d_{i,t} \tag{9}$$

$$f_5(s_t) = -\sum (d_{i,t} - p_{i,t} * x_{i,max}) \tag{10}$$

其中, P_t 表示 t 时段最长的停车时间; θ_1 、θ_2 分别代表算术序列权重和几何序列权重; f_1 、f_2 是基于目标函数而建立的特征函数,分别为充电站每时段向新进站车辆收取的费用和向电力公司支付的购电费用; f_3 、f_4 、f_5 都是用来约束所选择的策略, f_3 、f_4 是在文献[13]的基础上用来防止充电站做出影响未来收益的过激决策, f_5 是避免做出让车辆充电需求无法得到满足的决策。

引入特征函数后,Critic 部分价值函数的计算可由特征函数的线性组合近似得出 $v(s_t) = v(s_t, \omega) = \sum \omega_i f_i(s_t)$,Actor 部分策略函数同理可近似为 $\pi(a_t \mid s_t) = \pi_\theta(a_t \mid s_t) = \sum \theta_i f_i(s_t)$,有效减少了网络的输入维度。

3 算例分析

以某地区充电站信息采集系统实测数据为基础,取某年 1 月 1 日至 1 月 31 日的充电数据作为原始环境,测试集选取次年 1 月份的实测数据。假设充电站最高功率为 300 kW,对车辆进站情况和充电需求满足情况的观察间隔为 5 分钟,电价的变化情况为每小时变化一次。充电需求响应机制函数设置为 $D_i(r_t) = \beta_{i,1} r_t + \beta_{i,2}$,假设参与充电的汽车种类分为三种,对应的具体参数见表 1。

表 1 不同类型充电汽车的参数

参数	汽车 1	汽车 2	汽车 3
B_1/(kWh/ $)	-1	-4	-25
B_2/(kWh)	6	15	100
最大充电时间/min	30	120	720

为直观证明所提模型能较大程度提高充电站收益,引入以下算法对比:

Sarsa 算法:一种 value based 的强化学习算法,在模型未知的情况下,先利用当前策略 π 估算动作价值函数值 $Q(s,a)$,再通过 $Q(s,a)$ 值来更新策略,交替迭代得到最优策略和最优动作价值函数。为了避免陷入局部最优使用探索-利用机制,不断尝试新的动作[14]。

Policy Gradient 算法:一种 policy based 的强化学习算法,直接对策略进行近似表示,使用梯度上升法寻找最优策略。采用了类似 MC 算法的学习思路,需要遍历整个状态序列才能进行迭代[17]。

3.1 充电站收益

图 3 所展示的是充电站采用上述三种算法所获得平均收益的变化情况。在 3 000 轮迭代内,所提的 A3C 模型对提高充电站收益最明显,基本能稳定正收益;其次是 Sarsa 算法,充电站收益很少有正的大多为在 0 和-1 000 之间变化;最后是 PG 算法,虽然曲线收敛很快但没有正收益。

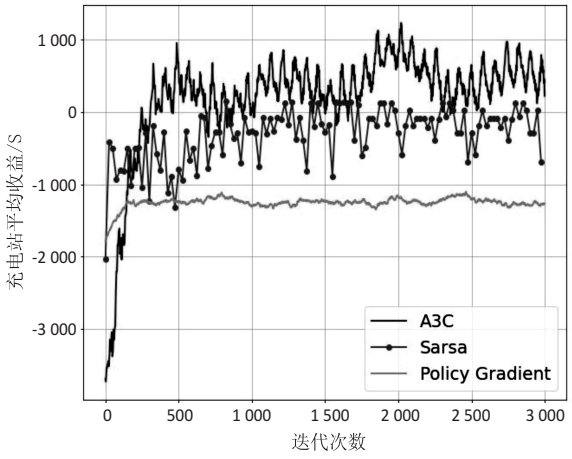

图 3 三种算法的平均收益

3.2 充电负荷方差

为验证所提模型能有效抑制充电负荷方差,将通过 A3C 算法与 Sarsa 算法训练得到的模型分别用于实际充电行为的调度。结果如图 4 所示,在连续 50 个小时的测试集上,不加任何调度完全无序充电得到的原始充电负荷波动较大,通过计算得出这 50 个小时内平

均充电负荷为 7.93 MW,进而得出负荷方差为 24.41 MW;利用 Sarsa 算法进行有序充电后的充电负荷波动幅度有所减少,通过计算相应的平均充电负荷最终可得负荷方差为 9.46 MW,与无序充电相比有了较大改善;结果最好的是提出的 A3C 算法,利用 A3C 的有序充电调度后充电负荷方差经计算可减少为 1.16 MW。

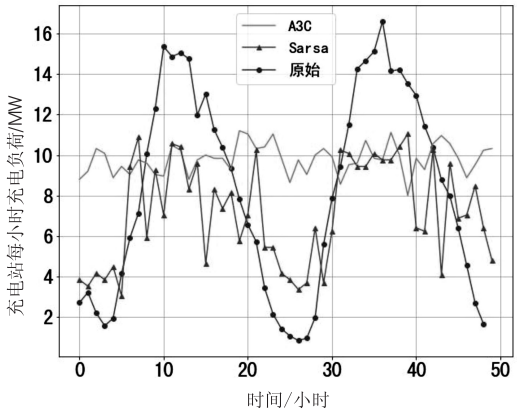

图 4　充电负荷方差

3.3　削峰填谷

比较典型日下电网 24 小时基础负荷与充电站 24 小时的充电功率,验证 A3C 算法能否实现削峰填谷。实验结果如图 5 所示,在电网基础负荷较高时,充电站会选择提高充电费用来鼓励用户不采取充电行为,相应的充电功率会降低防止电网峰值过高;同理,在电网基础负荷过低时,降低充电费用鼓励用户充电提高充电功率,提高电网谷值。该实验表明,所提模型除了能提高充电站收益外还有削峰填谷的作用。

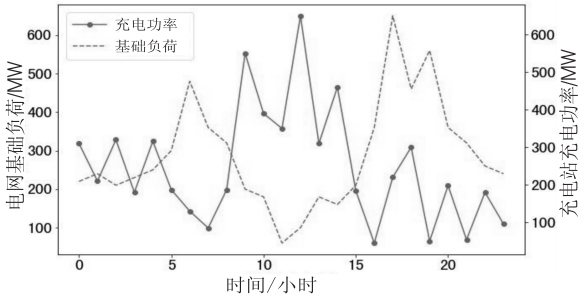

图 5　对电网充电负荷影响

4　结束语

提出一种无模型数据驱动的有序充电优化方法,不依赖任何先验模型和假设,把提高充电站收益和削峰填谷减少充电负荷方差作为目标,动态制定充电费用和充电功率。将充电问题抽象成 MDP 问题,利用深度强化学习算法 A3C 求解模型未知情况下的最优解。

为解决由于车辆进站的随机性造成输入维度的不确定性,Actor 和 Critic 都使用特征函数的线性组合来近似价值函数和策略函数。利用某地区充电站的实测数据进行仿真分析。结果表明,A3C 算法除了能提高充电站收益外,还能较大程度减少充电负荷方差,适应电网基础负荷的变化实现削峰填谷。

参考文献:

[1] 吴巨爱,薛禹胜,谢东亮,等. 电动汽车参与运行备用的能力评估及其仿真分析[J]. 电力系统自动化,2018,42(13):101-107.

[2] 赵俊华,文福拴,杨爱民,等. 电动汽车对电力系统的影响及其调度与控制问题[J]. 电力系统自动化,2011,35(14):2-10.

[3] JI Zhenya,HUANG Xueliang. Plug-in electric vehicle charging infrastructure deployment of China towards 2020:policies,methodologies,and challenges[J]. Renewable and Sustainable Energy Reviews,2018,90:710-727.

[4] YANG Jun,HE Lifu,FU Siyao. An improved PSO-based charging strategy of electric vehicles in electrical distribution grid[J]. Applied Energy,2014,128:82-92.

[5] 余晓玲,余晓婷,韩晓娟. 基于思维进化算法的电动汽车有序充电控制策略[J]. 电力工程技术,2017,36(6):58-62.

[6] SHAO Chengcheng,WANG Xifan,SHAHIDEHPOUR M,et al. Partial decomposition for distributed electric vehicle charging control considering electric power grid congestion[J]. IEEE Transactions on Smart Grid,2017,8(1):75-83.

[7] VANDAEL S,CLAESSENS B,ERNST D,et al. Reinforcement learning of heuristic EV fleet charging in a day-ahead electricity market[J]. IEEE Transactions on Smart Grid,2015,6(4):1795-1805.

[8] 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):4042-4048.

[9] SADEGHISNPOURHAMAMI N, DELEU J, DEVELDER C. Definition and evaluation of model-free coordination of electrical vehicle charging with reinforcement learning[J]. IEEE Transactions on Smart Grid,2020,11(1):203-214.

[10] ZHANG Hongcai,HU Zechun,MUNSING E,et al. Data-driven chance-constrained regulation capacity offering for distributed energy resources[J]. IEEE Transactions on Smart Grid,2018,10(3):2713-2725.

[11] AKHAVA-REZAI E,SHAABAN M F,EL-SAADANY E F,et al. Managing demand for plug-in electric vehicles in unbalanced LV systems with photovoltaics[J]. IEEE Transactions on Industrial Informatics,2017,13(3):1057-1067.

[12] GHAVAMI A,KAR K. Nonlinear pricing for social optimality of PEV charging under uncertain user preferences[C]// Proc of 48th annual conference on information sciences and systems. NJ:Princeton Press,2019:1-6.

[13] CHEKIRED D A,KHOUKHI L,MOUFTAH H T. Decentralized cloud-SDN architecture in smart grid:a dynamic pri-

(下转第 199 页)