DOI:10.12158/j.2096-3203.2021.01.026

基于时序差分学习的充电站有序充电方法

江明[1], 许庆强[1], 季振亚[2]
(1. 国网江苏省电力有限公司,江苏 南京 210024;
2. 南京师范大学电气与自动化工程学院,江苏 南京 210046)

摘 要:电动汽车有序充电是智能用电领域的重要议题。传统的模型驱动方法需对充电行为建模,但受相关参数的强随机性等影响,相关模型不能完全反映充电行为的不确定性。考虑到数据驱动下的无模型强化学习(MFRL)具有不依赖先验建模、适应强非线性关系样本数据的优势,提出将其应用于充电站的有序充电负荷优化。针对性地构建以用户充电需求能否获得满足为状态的马尔可夫决策过程(MDP),并利用充电完成度指标和满意度惩罚项改进代价函数。具体采用增量式的时序差分学习(TDL)算法训练历史数据,以保证数据规模下的计算性能。算例以充电站实测数据为环境,结果表明,在无需对充电行为进行先验建模的情况下,所提方法能够准确、快速地制定充电站有序充电计划。

关键词:电动汽车;有序充电;无模型强化学习;数据驱动方法;马尔可夫决策过程(MDP)
中图分类号:TM76 **文献标志码**:A **文章编号**:2096-3203(2021)01-0181-07

0 引言

近年来,电动汽车数量持续增加,对充电负荷进行有序调控将有助于缓解电动汽车给电网带来的峰谷差加剧、变压器过载等问题[1]。通过有序调控,大量充电负荷能以需求侧资源的形式直接或间接地参与电网调度体系[2]。为此,国内外学者提出了多种电动汽车充电调控结构,其中,充电站因其聚合控制的优势,能极大地降低网侧调度机构对大量电动汽车单体的关注,将成为充电调控的重要实施主体[3—4]。

充电站有序充电的研究主要包括建模预测和求解算法[5],预测以统计模型为主,如确定概率分布[6]、随机分布[7]、出行链模拟[8—9]等;算法则需综合车辆数量、抽样等对求解精度和速度的影响,包括拉格朗日松弛[10]、粒子群优化[11]、思维进化算法[12]等。但在实际情况下,充电需求参数的随机性强,耦合复杂[13],而充电站的充电设施数量有限,统计特征互补性差,模型驱动的有序充电方法面临性能瓶颈。

随着充电站运营,充电数据持续积累,数据驱动方法具有无需先验建模、适应复杂数据关系的优势,这为充电站的有序调控提供了新思路,其中,无模型强化学习(model-free reinforcement learning,MFRL)框架正日益受到重视。文献[14]提出利用MFRL 分析电价,获得经济性最优的充电策略;文献[15—16]以 Q-learning 算法调控充电站总功率;文献[17]提出一种基于蒙特卡洛学习(Monte Carlo learning,MCL)算法的充电站负荷充电预测算法;文献[18]以 MFRL 框架适应充电负荷的不确定性。然而,上述研究对各用户充电需求获得满足的考虑有所欠缺,且求解速度还有提升空间。

文中充分利用江苏某充电站采集积累的实测充电数据,提出一种基于时序差分学习(temporal difference learning,TDL)算法的有序充电优化方法。首先面向各用户的充电需求,针对性地设计以马尔可夫决策过程(Markov decision process,MDP)模拟的 MFRL 任务环境,并通过充电完成度指标和惩罚因子项完善代价函数,而 TDL 算法的增量式在线学习能力能够大幅提升性能。算例以充电站参与日前调度进行削峰填谷为例,根据实测数据,比较多种算法在准确度和求解速度上的性能。

收稿日期:2020-07-28;修回日期:2020-08-20
基金项目:江苏省自然科学青年基金资助项目(BK20190710)

1 基于 MDP 的有序充电建模

1.1 强化学习分类概述

强化学习任务以 MDP 为描述对象,MDP 由状态 X,动作 U,转移函数 P 和代价函数(或奖励函数) R 构成,共同组成环境。若环境中 X,U,P,R 各要素均已知,即对于任意状态 x,x' 和动作 u,其转移概率 $p(x'|x,u)$ 具有定量表达,则可以进行有模型的学习(model-based learning,MBL),这也是有序充电算法的常用技术。此时,学习器可以通过确定性或概率性的动态规划算法,以全概率展开的方式精确地

找到最优累积代价，确认最优策略。但实际应用时，电动汽车充电特性参数模型很难精确，转移概率往往缺少准确评估。此时，通过将有序充电问题看作免模型的强化学习过程，即 MFRL 框架，可以不再依赖先验和可用的转换和奖励模型，从而避免欠拟合、过拟合等现象对结果精确性的影响。

在 MFRL 框架中，最直接的策略评估替代算法为 MCL 算法，即通过多次采样并取平均累积代价的方式，将其作为期望累计代价的近似。这一算法能克服模型未知时的策略估计困境，但需等待整个采样轨迹的完成，是一类批处理式过程。TDL 算法在 MCL 的基础上进一步结合了动态规划原理，能够以增量处理方式在线更新估计。相比 MCL 算法，TDL 算法可以明显减少迭代最优方程的计算代价，效率更高。因此，TDL 算法用于电动汽车有序充电问题的求解时，有望在较为理想的时间内获得收敛。

上述强化学习的分类逻辑如图 1 所示。

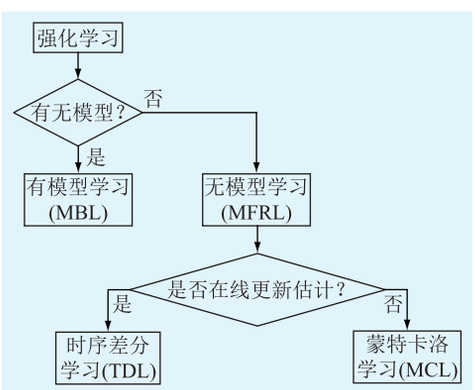

图 1　强化学习的主要算法分类

Fig.1　Classification of the main algorithms for reinforcement learning

1.2　有序充电的 MDP

目前，MDP 在有序充电领域正获得广泛应用，但不同于对电价[14]、总支出[15]、电压控制[16]、充电量与离网时刻关系[17]等环境的构造，文中侧重于将 MDP 描述为每个电动汽车用户的充电需求能否获得满足的决策框架。

考虑到有序充电的决策应保证用户充电需求获得满足，因此，不同于直接对充电负荷进行预测，MDP 需细化到电动汽车数量、入网和离网时刻、净充电量等多个充电行为参数。对电动汽车 i，具体参数包括：接入电网时刻 $T_{\mathrm{arr},i}$，离网时刻 $T_{\mathrm{dep},i}$，入网时的能量状态 $E_{\mathrm{arr},i}$，离网时的期望能量状态 $E_{\mathrm{dep},i}$ 及充电功率 P_{ch}。当考虑用户的充电需求时，对任意有序充电策略，要求使电动汽车 i 在 $T_{\mathrm{dep},i}$ 时刻离网时，其电量能够达到期望能量状态 $E_{\mathrm{dep},i}$。根据 MDP，对应的电动汽车充电行为四元组各要素的具体含义给定如下。

记入网电动汽车是否需要继续充电的状态空间为 X，根据用户的充电需求是否能被满足，X 包含 4 种状态，有：$X=\{x_0=$“可以灵活充电”，$x_1=$“必须立即充电”，$x_2=$“已满足期望电量而无需继续充电”，$x_3=$“期望电量难以完成”$\}$。对时刻 $h\in(T_{\mathrm{arr},i},T_{\mathrm{dep},i})$，电动汽车 i 所属状态 x_i 由下式决定：

$$x_i=\begin{cases}x_0 & E_{\mathrm{dep},i}-E_{i,h}<P_{\mathrm{ch}}(T_{\mathrm{dep},i}-h)\\ x_1 & E_{\mathrm{dep},i}-E_{i,h}=P_{\mathrm{ch}}(T_{\mathrm{dep},i}-h)\\ x_2 & E_{\mathrm{dep},i}-E_{i,h}=0\\ x_3 & E_{\mathrm{dep},i}-E_{i,h}>P_{\mathrm{ch}}(T_{\mathrm{dep},i}-h)\end{cases}\tag{1}$$

式中：$E_{i,h}$ 为电动汽车 i 在时刻 h 的能量状态。

由式(1)可以看出，定义的状态空间中，下一时刻的状态仅由当前状态决定，与历史状态无关，该过程符合 MDP 的性质。文中不考虑电动汽车向电网放电的情景，且认为充电功率均相等，即在给定的$[E_{\mathrm{arr},i},E_{\mathrm{dep},i}]$区间内充电功率恒定。

将有序充电决策的动作空间记为 U，包括不充电($u_0=0$)和充电($u_1=1$)两类动作，即：$U=\{u_0=0,u_1=1\}$。相应地，当在状态 $x\in X$ 下执行动作 $u\in U$ 时，根据转移函数，有 $p(x'\mid x,u)$ 的潜在概率会转移到另一个状态 $x'\in X$，伴随这一转移产生的代价值为 $r(x'\mid x,u)$。

电动汽车充电任务 MDP 如图 2 所示。

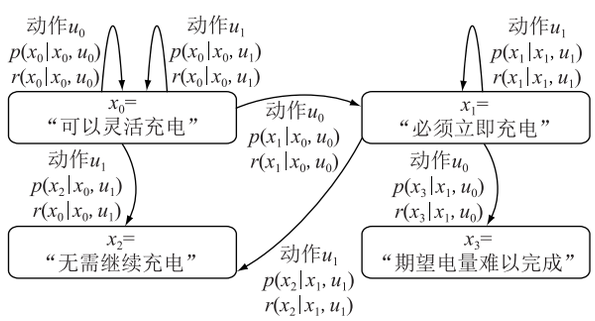

图 2　电动汽车充电任务 MDP

Fig.2　MDP for the electric vehicle charging task

当状态 $x_0=$“可以灵活充电”时，若动作选择 $u_0=0$，则下一状态会有一定的概率 $p(x_1\mid x_0,u)$ 转移到状态 $x_1=$“必须立即充电”，也会有一定的概率 $p(x_0\mid x_0,u)$ 保持当前状态不变，且这些动作返回的代价值较小；当状态为 $x_1=$“必须立即充电”时，若动作选择 $u_0=0$，则该状态会有概率 $p(x_3\mid x_1,u)$ 转移至 $x_3=$“期望电量难以完成”，而考虑到处于状态 x_3 时的用户充电需求将难以获得满足，且状态 x_3 难以恢复到其他状态，选择这一动作返回的代价值则相应较大。此外，要求每辆电动汽车初始设置时的

离网时刻与期望能量状态不能处于状态 x_3,避免初始状态对代价值的干扰。

1.3 有序充电策略决策的代价函数

将参与有序充电的电动汽车集群规模记为 I,时间周期记为 H。一般地,策略可以选择削峰填谷、追踪可再生能源出力、最小化电费等作为优化目标。为减少其他变量引入,文中将削峰填谷作为有序充电策略的优化目标,其值由累积充电负荷的二次项表达。当 $h\in H$,对电动汽车 $i\in I$ 执行充电动作 $u_{i,h}$ 时,策略的代价函数为:

$$\begin{cases} c_{\text{origin}}(\boldsymbol{x},\boldsymbol{u})=\left(\sum_{h=1}^{H}\sum_{i=1}^{I}u_{i,h}\right)^2 \\ u_{i,h}=\begin{cases}u_{i,h} & h\in(T_{\text{arr},i},T_{\text{dep},i})\\ 0 & h\notin(T_{\text{arr},i},T_{\text{dep},i})\end{cases}\end{cases} \tag{2}$$

式中:$\boldsymbol{x},\boldsymbol{u}$ 分别为电动汽车集群 I 在周期 H 内的状态空间 x 和动作空间 u 所构成的矩阵。

考虑到电动汽车 i 的用户充电需求应能够被满足,在调度周期内这一满意度的约束表示为:

$$E_{\text{arr},i}+u_{i,h}P_{\text{ch}}=ET_{\text{dep},i} \tag{3}$$

为使式(2)和式(3)能够更好地用于强化学习,首先定义一个“充电完成度”指标以减少用户充电特性参数的数量,再利用惩罚因子法将式(3)中的约束写入代价函数,以使强化学习的目标中包含满意度。具体步骤如下。

(1) 定义充电完成度指标。充电完成度 φ 用以表示电动汽车入网后充电量达到期望电量时所需的累积充电时长。电动汽车 i 的充电完成度 φ_i 与 $E_{\text{arr},i}$,$E_{\text{dep},i}$,P_{ch} 等参数的转换关系为:

$$\varphi_i=(E_{\text{dep},i}-E_{\text{arr},i})/P_{\text{ch}} \tag{4}$$

利用式(4)进行替代后,电动汽车充电行为特征相关的参数将从 $\{T_{\text{arr},i},T_{\text{dep},i},E_{\text{arr},i},E_{\text{dep},i},P_{\text{ch}}\}$ 减少到 $\{T_{\text{arr},i},T_{\text{dep},i},\varphi_i\}$,强化学习需处理的参数个数和数据量相应减少。

(2) 引入用户充电满意度惩罚项。不同于数学规划,强化学习难以直接处理约束,需要通过其他方法解决,如文献[19]中的温控负荷调度时,针对强化学习结果中不符合用户需求的控制操作,需要通过增加后备控制器来拒绝执行相关指令。文中将满意度约束直接引入代价函数,通过惩罚因子法,将原问题转化为非约束优化。

对于电动汽车充电行为,违反满意度约束的情况即电动汽车 i 充入电量未达到或超过所需电量,在定义的充电完成度与动作空间下,惩罚项为:

$$\xi_i=\left|\varphi_i-\sum_{h=1}^{H}u_{i,h}\right| \tag{5}$$

利用惩罚因子法实现约束的原理在于,若对某辆电动汽车在某时刻采取不充电动作(即“$u_0=0$”)而导致该车充电量无法获得满足(即状态转移至 $x_3=$“期望电量难以完成”)的情况,则将付出极大的代价值,促使策略做出不违反相关约束的选择。惩罚因子法通过一个较大的惩罚因子 M 实现,M 是一个人工选择的参量,应满足:

$$M+(I-1)^2>I^2 \tag{6}$$

改进后的代价函数为:

$$c(\boldsymbol{x},\boldsymbol{u})=\left(\sum_{h=1}^{H}\sum_{i=1}^{I}u_{i,h}\right)^2+M\sum_{i=1}^{I}\xi_i \tag{7}$$

2 基于 TDL 算法的有序充电策略求解

2.1 有序充电策略的状态-动作值函数

强化学习的任务是通过在环境中不断尝试,模拟不同的策略和相应的代价值,习得一个最优策略 $\boldsymbol{\pi}^*$,使执行该策略后的累积代价最小。将电动汽车有序充电问题看作强化学习时,根据常用的 T 步强化学习方案,对任意有序充电策略 $\boldsymbol{\pi}$,其代价由累积代价函数值决定:

$$J_{\pi,T}(\boldsymbol{x})=E\left[\frac{1}{T}\sum_{t=1}^{T}c_t(\boldsymbol{x},\boldsymbol{u})\,|\,\boldsymbol{x}_0=\boldsymbol{x},\boldsymbol{u}_0=\boldsymbol{u}\right] \tag{8}$$

式中:$\boldsymbol{x}_0,\boldsymbol{u}_0$ 分别为电动汽车集群充电任务和充电策略的起始状态矩阵;$E[\cdot]$ 为期望。

强化学习将对某个状态 $\boldsymbol{x}$ 上执行动作 $\boldsymbol{u}$ 后获得的值表示为状态-动作值函数,记为 $Q_{\pi,T}(\boldsymbol{x},\boldsymbol{u})$,这里用 T 步累积代价表示:

$$Q_{\pi,T}(\boldsymbol{x},\boldsymbol{u})=E[c_t(\boldsymbol{x},\boldsymbol{u})+J_{\pi,T}(\boldsymbol{x}')] \tag{9}$$

式中:$\boldsymbol{x}'$ 为前一次在状态 $\boldsymbol{x}$ 执行动作 $\boldsymbol{u}$ 后转移到的状态。

记最优策略 $\boldsymbol{\pi}^*$ 下的最优累积代价为 $Q^*(\boldsymbol{x},\boldsymbol{u})$,其值必能满足最优贝尔曼方程:

$$Q^*(\boldsymbol{x},\boldsymbol{u})=E[c_t(\boldsymbol{x},\boldsymbol{u})+\min Q^*(\boldsymbol{x}',\boldsymbol{u})] \tag{10}$$

考虑到强化学习过程最终获得的最优累计代价仅为近似值,而非严格证明的理论最优值,以 $\bar{Q}^*$ 表示,此时,近似最小化累积代价及其对应的最优策略 $\boldsymbol{\pi}^*$ 满足:

$$\boldsymbol{\pi}^*=\arg\min_{\boldsymbol{\pi}}\bar{Q}^*(\boldsymbol{x},\boldsymbol{u}) \tag{11}$$

最优策略 $\boldsymbol{\pi}^*$ 即为有序充电策略最终选择执行的动作:

$$\boldsymbol{u}=\boldsymbol{\pi}^* \tag{12}$$

2.2 基于 TDL 算法的求解过程

在 MFRL 框架下,因充电 MDP 任务中各状态间

的转移概率未知，策略评估只能通过执行选择的动作来观察转移的状态和相应的代价值。一种直接的策略评估替代方式是 MCL 算法，但其需在完整的采样轨迹完成后更新所有的状态-动作对；另一种方法是，若基于前 t 个采样已估计的值函数 $Q_{\pi,t}(\boldsymbol{x},\boldsymbol{u})$，对其进行增量式更新，可以在每步之后仅增加 1 个增量，不仅提高了求解效率，而且不影响 Q_t 为累积代价之和的性质。

TDL 算法中，基于前 t 个采样得到 $Q_{\pi,t}(\boldsymbol{x},\boldsymbol{u})$、式(8)和式(9)，得：

$$J_{\pi,t}(\boldsymbol{x}) = E\left[\frac{1}{t}\sum_{1}^{t} c_t(\boldsymbol{x},\boldsymbol{u})\right] \quad t \in T \tag{13}$$

$$Q_{\pi,t+1}(\boldsymbol{x},\boldsymbol{u}) = Q_{\pi,t}(\boldsymbol{x},\boldsymbol{u}) + \frac{1}{\alpha}[J_{\pi,t}(\boldsymbol{x}) - Q_{\pi,t}(\boldsymbol{x},\boldsymbol{u})] \tag{14}$$

式中：α 为更新步长，是一个较小的正系数。

下一步仅需给 $Q_{\pi,t}(\boldsymbol{x},\boldsymbol{u})$ 加上式(14)第二行的增量，通过增量求和，有：

$$Q_{\pi,t+1}(\boldsymbol{x},\boldsymbol{u}) \leftarrow Q_{\pi,t}(\boldsymbol{x},\boldsymbol{u}) + \frac{1}{\alpha}[c(\boldsymbol{x}'|\boldsymbol{x},\boldsymbol{u}) + \gamma Q_{\pi,t}(\boldsymbol{x}',\boldsymbol{u}') - Q_{\pi,t}(\boldsymbol{x},\boldsymbol{u})] \tag{15}$$

式中：γ 为代价值的折扣因子；$\boldsymbol{u}'$ 为策略 π 在 $\boldsymbol{x}'$ 上选择执行的动作。

针对强化学习的探索-利用困境，利用 ε-贪心策略作为克服该困境平衡规则[20]。基于概率对两者折中，即在 $(t-1)$ 步后，平均代价 $Q_t(\boldsymbol{x},\boldsymbol{u})$ 更新为：

$$Q_{t+1}(\boldsymbol{x},\boldsymbol{u}) \leftarrow \frac{1}{t}[(t-1)Q_t(\boldsymbol{x},\boldsymbol{u}) + c_t(\boldsymbol{x},\boldsymbol{u})] \tag{16}$$

所提电动汽车有序充电方法中 TDL 算法的迭代过程主要包括以下步骤：(1) 初始化，包括定义环境，动作空间 U，起始状态 $\boldsymbol{x}_0$，起始策略 π_0，探索概率 ε 及步数 T；(2) 策略在 $t=1$ 至 T 中按步执行并更新，包括按 ε-贪心策略执行策略 π 中对应的各动作 u，求得 $c(\boldsymbol{x}'|\boldsymbol{x},\boldsymbol{u})$ 与转移到的状态 $\boldsymbol{x}'$，按式(16)更新平均代价，并求得对应的动作 $\boldsymbol{u}'$；(3) 当策略下的最优累积代价满足最优贝尔曼方程时，结束迭代，并输出最优策略 π^*。TDL 算法的收敛性证明可见文献[21]。

3 算例分析

3.1 算例设置

以江苏某充电站信息采集系统实测数据为基础，取某年 8 月至次年 7 月共 11 个月数据为原始环境，测试日期为次年 8 月 $d=\{1,2,\cdots,31\}\in D$，测试日期中每日的原始充电负荷曲线分布如图 3 所示。其中，以充电累积负荷中位数对应日期作为典型日($d=5$)，阴影区间表示该测试月中所有日期下对应时段负荷的分布范围。

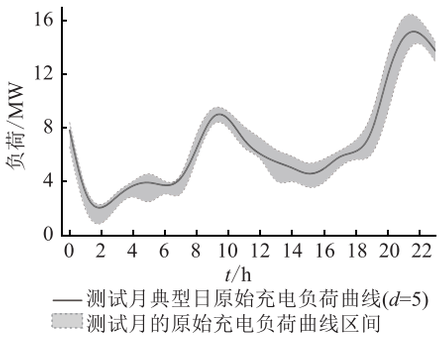

图 3 测试日期的原始充电负荷曲线分布

Fig.3 Distribution of the original charging load curves in the simulation days

测试前，根据 $E_{\mathrm{arr},i}$，$E_{\mathrm{dep},i}$，P_{ch} 求得 φ_i，$T_{\mathrm{arr},i}$，$T_{\mathrm{dep},i}$，φ_i，均按 1 h 的时间间隔向上取整。该准备适用于所有算法，且执行时间短，不计入时间。假设电动汽车和充电设施均支持有序充电，选择日前调度计划作为考虑对象，并将日前预测的有序充电结果称为有序充电日前计划，以区别于测试日当天的实际有序充电曲线。

为直观比较，引入以下算法：

(1) 基准最优(best of performance，BOP)算法。根据测试日的充电实测数据，通过式(2)和式(3)直接求解。这一方法获得的是理想最优结果，属于后验结果，实践中无法达到。

(2) MBL 算法。先根据环境中数据，分别对每辆电动汽车的 $\{T_{\mathrm{arr},i},T_{\mathrm{dep},i},\varphi_i\}$ 拟合建模，再以动态规划算法求解。

(3) MCL 算法。与 TDL 算法相比，MCL 法具体步骤中不是按步执行更新，而是对完整的各步结果的总和进行平均。

策略执行设置从第 d 日 0 h 开始，每轮迭代所赋起始状态初值均由当日即时接入状态决定，且为保证独立性，每日决策周期结束后重新学习。

编程语言为 Python，平台为 Jupyter Notebook，在处理器 Intel i7、主频 3.4 GHz 的计算机上执行。

3.2 准确性分析

首先，以典型日为例进行比较，典型日总充电量为 163.6 MW·h，调度前负荷峰谷差为 13.56 MW。直接观察 4 种算法下的有序充电日前计划曲线，如图 4 所示。考虑到优化目标是负荷波动平抑，曲线越平缓则表明平抑效果越好。总体上，与优化前的原始充电负荷曲线相比，4 种算法的峰值降低，谷值提高，日前计划曲线均更为平缓。

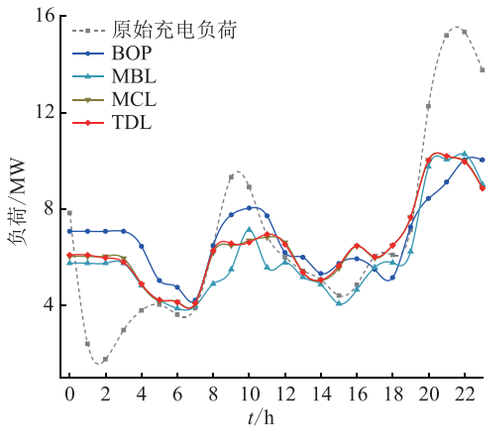

图 4 不同算法下典型日的有序充电负荷曲线

Fig.4 Coordinated charging load curves in the typical day under different algorithms

为定量比较各曲线代表的有序充电计划效果,表 1 给出了图 4 中 24 h 总充电量和充电负荷峰谷差的数值分析。根据表 1 可知,以 BOP 算法为基准,MBL、MCL 和 TDL 算法下的总充电量在日前计划时分别有 -11.7%,-4.7%,-4.5% 的偏差,即 MFRL 框架下的 MCL 和 TDL 算法对当日总用电量的估计更准确。与此同时,负荷峰谷差的偏差分别为 9.97%,4.81%,4.81%,平均绝对误差分别为 53.3 MW,40.2 MW,39.6 MW。因此,MCL 和 TDL 算法相比 MBL 算法更接近 BOP 算法中的理论最优值。以上表明,就典型日而言,文中所提 MFRL 框架下的 MCL 算法和 TDL 算法均优于 MBL 算法,对结果准确性的影响较小。

表 1 不同算法下典型日的有序充电计划定量分析

Table 1 Quantitative analysis of coordinated charging plans of different algorithms for the typical day

参数	算法结果			
	BOP	MBL	MCL	TDL
计划总电量/(MW·h)	163.6	144.4	156.0	156.1
计划总电量偏差/%		-11.7	-4.7	-4.5
负荷峰谷差/MW	9.04	6.4	6.1	6.1
负荷峰谷差偏差/%		9.97	4.81	4.81
平均绝对误差/MW		53.3	40.2	39.6

其次,考察全部测试日中各算法的准确性。为了便于观察,以 BOP 算法为基准,定义准确率 α_d 以标记其他算法的准确性。记 $P_{\mathrm{BOP},h}$ 为 BOP 算法下第 h 小时的充电负荷,对第 d 日中各小时,有:

$$\alpha_d \left| \sum_{h=1}^{H} P_{\mathrm{BOP},h} - P_{\mathrm{ch}} \sum_{h=1}^{H} \sum_{i=1}^{I} u_{i,h} \right| \Big/ \sum_{h=1}^{H} P_{\mathrm{BOP},h} \quad (17)$$

图 5 统计了 4 种算法下准确率数值分布。可以看出,MBL 算法尽管可能在部分日期中较优(如对应散点所示),但在总体分布上仅近似甚至劣于 TDL 与 MCL 算法,其计划能否取得精确的结果具有更强的不稳定性(如对应区间所示),这是由于同一组充电行为特征模型在不同日期的准确性是难以控制的。而对于 MFRL 框架下的 MCL 与 TDL 算法,两者的准确率在数值分布上相对较为接近,即结果的准确性近似。

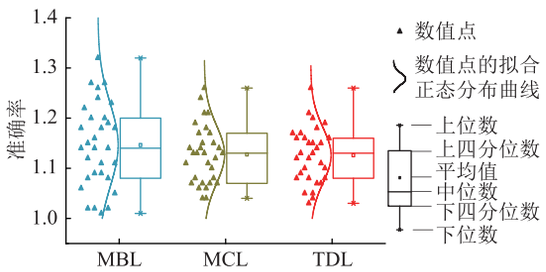

图 5 所有测试日有序充电计划的准确率分布

Fig.5 Distribution of the accuracy rates of coordinated scheduling results over all the simulation days

因此,在 MFRL 框架下,有序充电策略的求解不仅无需进行先验建模,而且相对于基于某个固定统计模型的 MBL 算法,能够更稳定地获得更接近最优效果的曲线。而相比 MCL 算法,文中所用的增量式在线更新 TDL 算法在计划的准确性方面基本未受影响。

3.3 求解速度分析

继续观察各算法执行时的求解速度。首先,观察强化学习下各算法的收敛过程。MBL、MCL 和 TDL 算法下,各测试日求解时长的中位数所在日期分别为 8 月 7 日、23 日和 18 日,对应的收敛曲线如图 6 所示。有模型的 MBL 算法提供了较优的求解速度,然而,尽管 TDL 和 MCL 算法收敛速度显著慢于 MBL 算法,但 TDL 算法对应曲线的曲率大于 MCL 算法,收敛过程较快,从而可以更快地逼近最优解。

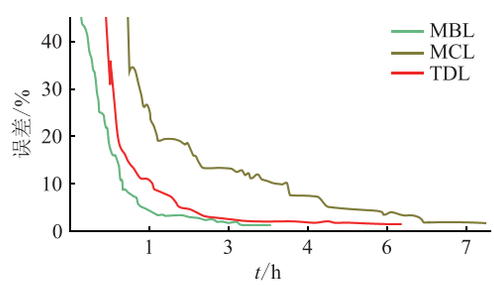

图 6 测试日求解时长中位数日期的收敛曲线

Fig.6 Convergence curves for the median value of solving time over all the simulation days

其次,考察所提算法的求解速度能否稳定在一个合理时长区间。考虑到日前市场周期 $H=24$ h,一般地,日前市场的出清规划要求不晚于前一天的

12 h 提交次日计划，因此，以 12 h 为求解建议时长的上限进行考察。图 7 刻画了 MCL 和 TDL 算法下所有测试日期的求解时长分布。对应该时长上限，TDL 算法未出现求解超时的情况，且相对集中，均明显少于 12 h，即求解时长较少，求解用时的分布比较集中。与此同时，MCL 算法下的各调度日期求解用时分布明显偏右，甚至会出现超过 12 h 的极端情况。

因此，在策略求解的快速性方面，尽管 MFRL 框架下的 MCL 算法和 TDL 算法相比 MBL 算法，确实出现了收敛相对较慢的问题，但 TDL 算法的求解速度明显优于 MCL 算法，能够满足日前调度对计算时长的要求。

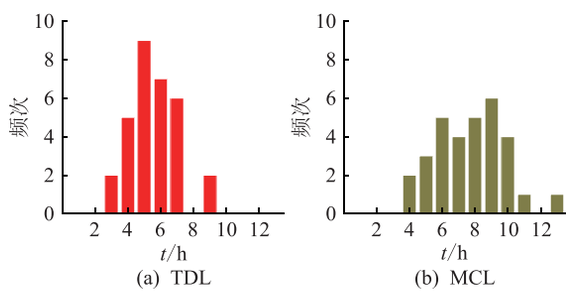

图 7 TDL 与 MCL 的收敛时间直方图

Fig.7 Histograms for the convergence time distributions of TDL and MCL

4 结论

文中提出一种无需先验建模的充电站有序充电优化方法，并基于江苏某充电站的实测数据进行仿真分析。主要结论如下：

（1）区别于依赖模型驱动的 MBL 算法，更依赖数据驱动的 MBFL 框架能够使有序充电策略不再依赖充电行为的精确先验模型，减少了步骤，精确性更高；

（2）MBFL 框架下，相比于传统的 MCL 算法，增量式在线更新的 TDL 算法可以显著提高收敛速度，更能适应日前计划对求解时间的要求；

（3）算例结果表明，实际运营中的充电站能够在满足用户充电需求的前提下，为电网提供有益的削峰填谷潜力。

文中提出的相关算法能够利用充电站采集积累的充电数据，直接进行有序充电计划，具备良好的工程实施能力。后续可以通过引入多分类器等算法继续提升求解效率，以及引入市场电价、可再生能源出力等其他数据，拓展到更多的优化场景。

本文得到国网江苏省电力有限公司科技项目（J2019016），江苏省高等学校自然科学研究面上项目（19KJD470004）资助，谨此致谢！

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.
WU Juai，XUE Yusheng，XIE Dongliang，et al. Evaluation and simulation analysis of reserve capability for electric vehicles[J]. Automation of Electric Power Systems，2018，42(13)：101-107.

[2] 占恺峤，胡泽春，宋永华，等. 含新能源接入的电动汽车有序充电分层控制策略[J]. 电网技术，2016，40(12)：3689-3695.
ZHAN Kaiqiao，HU Zechun，SONG Yonghua，et al. Electric vehicle coordinated charging hierarchical control strategy considering renewable energy generation integration[J]. Power System Technology，2016，40(12)：3689-3695.

[3] 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.
ZHAO Junhua，WEN Fushuan，YANG Aimin，et al. Impacts of electric vehicles on power systems as well as the associated dispatching and control problem[J]. Automation of Electric Power Systems，2011，35(14)：2-10.

[4] JI Z，HUANG X. 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.

[5] 王锡凡，邵成成，王秀丽，等. 电动汽车充电负荷与调度控制策略综述[J]. 中国电机工程学报，2013，33(1)：1-10.
WANG Xifan，SHAO Chengcheng，WANG Xiuli，et al. Survey of electric vehicle charging load and dispatch control strategies[J]. Proceedings of the CSEE，2013，33(1)：1-10.

[6] 李丹奇，郑建勇，史明明，等. 电动汽车充电负荷时空分布预测[J]. 电力工程技术，2019，38(1)：75-83.
LI Danqi，ZHENG Jianyong，SHI Mingming，et al. Prediction of time and space distribution of electric vehicle charging load[J]. Electric Power Engineering Technology，2019，38(1)：75-83.

[7] HASHEMI B，SHAHABI M，TEIMOURZADEH-BABOLI P. Stochastic-based optimal charging strategy for plug-in electric vehicles aggregator under incentive and regulatory policies of DSO[J]. IEEE Transactions on Vehicular Technology，2019，68(4)：3234-3245.

[8] 李琳玮，宁光涛，俞悦，等. 基于交通信息的多类型电动汽车综合充电需求研究[J]. 电力工程技术，2020，39(1)：191-199.
LI Linwei，NING Guangtao，YU Yue，et al. Comprehensive charging demand of multi-type electric vehicles based on traffic information[J]. Electric Power Engineering Technology，2020，39(1)：191-199.

[9] 李含玉，杜兆斌，陈丽丹，等. 基于出行模拟的电动汽车充电负荷预测模型及 V2G 评估[J]. 电力系统自动化，2019，43(21)：88-96.
LI Hanyu，DU Zhaobin，CHEN Lidan，et al. Trip simulation based charging load forecasting model and vehicle-to-grid evalu-

ation of electric vehicles[J]. Automation of Electric Power Systems,2019,43(21):88-96.

[10] SHAO C,WANG X,SHAHIDEHPOUR M,et al. Partial decomposition for distributed electric vehicle charging control considering electric power grid congestion[J]. IEEE Transactions on Smart Grid,2016,8(1):75-83.

[11] YANG J,HE L,FU S. An improved PSO-based charging strategy of electric vehicles in electrical distribution grid[J]. Applied Energy,2014(128):82-92.

[12] 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.
YU Xiaoling,YU Xiaoting,HAN Xiaojuan. Coordinated charging strategy for PEV charging stations based on mind evolutionary algorithm[J]. Electric Power Engineering Technology,2017,36(6):58-62.

[13] SADEGHIANPOURHAMAMI N,REFA N,STROBBE M,et al. Quantitive analysis of electric vehicle flexibility:a data-driven approach[J]. International Journal of Electrical Power & Energy Systems,2018,95:451-462.

[14] 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.

[15] 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.

[16] 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.
DU Mingqiu,LI Yan,WANG Biao,et al. Deep reinforcement learning optimization method for charging control of electric vehicles[J]. Proceeding of the CSEE,2019,39(14):4042-4048.

[17] SADEGHIANPOURHAMAMI 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.

[18] ZHANG H,HU Z,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.

[19] CLAESSENS B J,VRANCX P,RUELENS F. Convolutional neural networks for automatic state-time feature extraction in reinforcement learning applied to residential load control[J]. IEEE Transactions on Smart Grid,2016,9(4):3259-3269.

[20] CHEN X,GAO Y,WANG R. Online selective kernel-based temporal difference learning[J]. IEEE Transactions on Neural Networks and Learning Systems,2013,24(12):1944-1956.

[21] VAN SEIJEN H,MAHMOOD A R,PILARSKI P M,et al. True online temporal-difference learning[J]. The Journal of Machine Learning Research,2016,17(1):5057-5096.

作者简介:


江明

江明(1987),男,学士,工程师,从事电动汽车充电站的智能化管理、电力营销相关工作(E-mail:13505161772@126.com);

许庆强(1976),男,博士,教授级高级工程师,从事电动汽车充电站的智能化管理、电力营销相关工作;

季振亚(1988),女,博士,讲师,研究方向为电动汽车与电网互动、电力市场机制。

Coordinated charging approach for charging stations based on temporal difference learning

JIANG Ming[1], XU Qingqiang[1], JI Zhenya[2]

(1. State Grid Jiangsu Electric Power Co.,Ltd.,Nanjing 210024,China;
2. School of Electrical and Automation Engineering,Nanjing Normal University,Nanjing 210046,China)

Abstract:Coordinated charging of electric vehicles (EVs) is becoming an important topic for the smart demand management. Traditional model-driven methods are highly dependent on the accuracy of models for charging behavioral characteristics. However,affected by the strong stochastics of related parameters,etc.,the selection of relevant models cannot fully reflect their uncertainties. Considering that the data-driven model-free reinforcement learning algorithms has the advantages of not relying on pre-modeling,and adapting to data samples with strong nonlinear relationships,it is proposed to be applied to optimize the charging loads of the EV charging stations. In the Markov decision process customized for the satisfaction of EV charging need,both a charging completion degree index and a penalty term for user's charging satisfaction are introduced to improve the policy evaluating function. Specifically,in order to guarantee the computational speed underneath the volume of charging data,the temporal difference learning algorithm is used for the training with incremental updates. The simulation is carried out with the real-world data from one charging station. Results show that the proposed algorithm can accurately and quickly calculate the coordinated charging schedules without the pre-modeling for the EV charging behavior parameters.

Keywords:electric vehicle;coordinated charging;model-free reinforcement learning;data-driven approach;Markov decision process(MDP)

(编辑 钱悦)