

# 基于深度强化学习的电动汽车协调充电算法

张子霖

(中国科学技术大学 自动化系, 安徽 合肥 230026)

**摘要:** 针对具有多个目标的多辆电动汽车协调充电问题, 包括降低用户的充电成本并保证足够的电池电量, 同时避免变压器过载, 提出一种基于多智能体深度强化学习的分布式算法。该算法首先将上述问题建模为马尔可夫决策过程, 然后采用一种通信神经网络模型进行信息交互, 实现了全局信息在多个智能体上的分布式计算。最后, 通过仿真验证了该算法在住宅充电区域的有效性和可扩展性。

**关键词:** 电动汽车; 多智能体系统; 深度强化学习; 分布式充电调度

中图分类号: TP18

文献标识码: A

DOI: 10.19358/j.issn.2096-5133.2022.04.013

引用格式: 张子霖. 基于深度强化学习的电动汽车协调充电算法[J]. 信息技术与网络安全, 2022, 41(4): 83-89.

## A deep RL-based algorithm for coordinated charging of electric vehicles

Zhang Zilin

(Department of Automation, University of Science and Technology of China, Hefei 230026, China)

**Abstract:** A distributed algorithm based on multi-agent deep reinforcement learning (DRL) is proposed in this paper to solve the problem of coordinated charging of multiple electric vehicles (EVs) with multiple objectives, including reducing the charging cost, ensuring sufficient battery state of charge (SoC) for users, and avoiding the transformer overload. To this end, the above problem is formulated as a Markov Decision Process (MDP), and then a communication neural network (CommNet) model is adopted for information interaction to realize the distributed computation of global information on multiple EVs. Finally, simulations are presented to verify the effectiveness and scalability of the proposed algorithm in a residential charging area.

**Key words:** electric vehicle; multi-agent system; deep reinforcement learning; distributed charging scheduling

### 0 引言

电动汽车作为一种新型的分布式移动资源, 已成为智能电网发展的重要组成部分<sup>[1]</sup>, 越来越多的研究人员开始关注电动汽车在电网中的社会价值。但电动汽车的日益普及也不可避免地给用户和电网带来了一系列的问题和挑战。例如, 大量电动汽车的不协调充电, 将导致电网承受不可预测的负荷。因此, 如何设计多辆电动汽车在整个工作过程中的协调充电策略, 以降低能源成本, 保证用户对电动汽车荷电状态的满意度, 并将对配电网的影响降到最低, 是亟需解决的问题。

近年来, 为了解决电动汽车的协调充电问题, 出现了许多分布式调度方法。例如, 为了最大限度地提高电动汽车用户的便利性, 开发了一种分散的基于交替方向乘法器的优化算法<sup>[2]</sup>。为使电动汽车

充电成本最小化, 使用电动汽车充电概率模型建立了一个严格凸分散系统形式的多人博弈<sup>[3]</sup>。但上述研究只关注了单一的目标, 这在实际应用中往往是受限制的。针对多目标的协调充电问题, 通过多智能体自私协同优化的方案, 除了实现用户利益以外, 也保证避免变压器过载<sup>[4]</sup>, 但它没有充分考虑状态空间泛化和值函数逼近的影响, 导致拟合性能较差, 计算开销也不理想。针对该问题, 文献[5]考虑采用基于机器学习的方法, 其以收敛速度快、计算效率高而被广泛应用于大规模数据的高维问题中。

事实上, 基于强化学习的数据驱动方法作为机器学习的一个重要分支, 已经广泛应用于能源领域, 特别是电动汽车充电问题<sup>[6-8]</sup>。这是因为基于强化学习的方法可以在缺乏明确的数学模型的情况下模拟顺序决策问题, 并进一步获得对环境的精确

响应<sup>[9]</sup>。具体来说,通过建立具有可扩展状态表示的马尔可夫决策过程,以直接决策联合控制整个电动汽车集合<sup>[6]</sup>。此外,文献[7]考虑到电价的随机性和电动汽车的通勤行为,提出了一种结合电价预测的深度强化学习解决方案来优化实时调度。但以上两种基于强化学习的方法是集中实现的,显然不可能处理大规模动态数据的在线计算。针对电动汽车协调充电的分布式强化学习方法,文献[8]引入了一种分散的基于市场的方法,使多辆电动汽车能够积极参与电力市场。但其中采用的Q-learning是基于表格的方法,随着问题规模的扩大,会受到维数灾难的影响。因此,有必要寻求一种分布式强化学习算法,在保证问题规模可扩展和计算开销可接受的前提下,解决多目标的电动汽车协调充电问题。

基于上述研究,本文提出了一种基于深度强化学习的电动汽车协调充电分布式算法,重点关注降低用户的充电成本,确保日常行驶中足够的电池荷电状态,并且避免变压器过载。

## 1 问题描述

考虑一个居民区,当地的配电网连接到一定数目的家庭,每家都有一辆私人使用的电动汽车。在本文研究的协调充电问题中,所有参与信息交互的电动汽车之间存在通信链路,从而形成一个全连通的集合 $B=\{1,2,\dots,N\}$ 。整个时间周期为 $G=\{1,2,\dots,T\}$ ,即时间间隔为1h。在时间步 $t\in G$ ,电动汽车 $n\in B$ 获取本地状态信息(即电动汽车的位置和电池荷电状态)和电价信息,同时接收变压器的负载信息和相邻电动汽车的状态信息。然后每辆电动汽车根据上述信息进行局部决策,执行充电或不充电的具体行为。

为了电动汽车用户和电网方的利益,本文研究的电动汽车协调充电问题的目标是找寻最优的充电策略,减少电动汽车的充电总成本,确保足够的电池荷电状态,并避免变压器过载。需要注意的是,用户的通勤行为是不确定的,无法精确建模,因此可将电动汽车的充电行为看成一个随机过程。由于当前的信息足以预测未来的状态,故上述随机过程独立于历史,符合马尔可夫特性。基于以上讨论,将本文研究的多目标协调充电问题表示成一个包含五元素 $(S,A,P,R,\gamma)$ 的马尔可夫决策过程,以描述不确定性。其中 $S$ 是状态空间, $A$ 是所有可行的行为, $P$ 是状态转移概率, $R$ 是奖励函数, $\gamma$ 是折扣因

子。各元素的详细介绍如下:

(1)状态空间:记为 $S_t=\{s_t^1,s_t^2,\dots,s_t^N\}$ ,其中向量 $s_t^n=[E_t^n,L_t^n,U_t^n,p_{t-23}^n,p_{t-22}^n,\dots,p_t^n]$ 表示电动汽车 $n$ 在 $t$ 时的状态,它包含了以下四种信息: $E_t^n\in(0,1)$ 表示电动汽车 $n$ 在 $t$ 时的电池荷电状态(即当前剩余电池容量与电池全容量的比值),可能会受到位置和行为的影响。若电动汽车充电, $E_t^n$ 将增加20%,若其出行则 $E_t^n$ 减少20%<sup>[4]</sup>。 $L_t^n\in(0,1)$ 表示在 $t$ 时刻,电动汽车 $n$ 接收到的变压器负载占变压器最大期望负载的百分比。 $U_t^n=\{0,1\}$ 表示电动汽车 $n$ 在 $t$ 时的位置信息,0表示在家,1表示出行在外。 $P_t^n=(p_{t-23}^n,p_{t-22}^n,\dots,p_t^n)$ 表示电动汽车 $n$ 依次在 $t-23,t-22,\dots,t$ 时刻获取到的电价(即过去24h的电价)。

(2)行为空间:离散行为空间 $A_t=\{a_t^1,a_t^2,\dots,a_t^N\}$ 包含了所有可能的行为。 $a_t^n=\{0,1\}$ 表示电动汽车 $n$ 在状态 $s_t^n$ 下执行的行为,1表示充电,0表示不充电。

(3)奖励空间:定义为 $R_t=\{r_t^1,r_t^2,\dots,r_t^N\}$ ,其中 $r_t^n=r_{e,t}^n+r_{p,t}^n+r_{l,t}^n$ 表示电动汽车 $n$ 在 $t$ 时获得的奖励。它由三部分组成,每部分都对应着一个既定目标: $r_{e,t}^n=10*E_t^n$ ,表示 $t$ 时电动汽车 $n$ 用户对电池荷电状态满意度的奖励; $r_{p,t}^n=10/(p_t^n*c_t^n)$ 为能源成本减少程度的奖励,其中 $c_t^n$ 为电动汽车 $n$ 在 $t$ 时刻消耗的电量; $r_{l,t}^n=\{0,10\}$ 是为了避免变压器过载而设计的,当电动汽车 $n$ 接收到的变压器处于过载状态时,没有奖励。当变压器负载处于期望范围内,电动汽车 $n$ 用户得到的奖励为10。

在本文研究的多智能体系统中,参数化的策略函数为:

$$\pi_\theta(S,A)=P[A|S,\theta] \quad (1)$$

其中 $P$ 为通过参数 $\theta$ 将状态 $S$ 映射到行为 $A$ 上的概率分布函数。最后,根据在 $S_k$ 状态下执行 $A_k$ 后的期望回报来评估充电调度的质量:

$$Q^{\pi_\theta}(S_t,A_t)=E_{\pi_\theta}\left[\sum_{k=t}^T\gamma^{(k-t)}R(S_k,A_k)\right] \quad (2)$$

其中 $Q^{\pi_\theta}(S_t,A_t)$ 表示真实的行为值函数。接下来需要在状态转移概率 $P(S'|S,A)$ 未知的情况下,求解

上述马尔可夫决策过程。考虑到不能通过直接求解贝尔曼方程的方式来获得最优解,而无模型的深度强化学习算法可以保证智能体在与环境进行多次交互后获得最优策略,故本文开发了一个深度强化学习算法来求解所制定的马尔可夫决策过程问题。

## 2 策略方法

本节提出了一种基于神经网络模型的深度强化学习算法,来解决马尔可夫决策过程形式下的多目标电动汽车协调充电问题,其中利用通信神经网络(Communication Neural Network, CommNet)模型实现多辆电动汽车上的分布式计算。在后续讨论和分析中,智能体指的是电动汽车,环境指的是与电动汽车交互的住宅充电场景,并设定  $T=24$ 。

CommNet 是一种神经网络模型,常与多智能体强化学习结合使用。它可以促成多个智能体之间的合作,使它们根据各自的可观察状态采取行动之前进行信息交互。本文利用 CommNet 来帮助连接到局部电网的电动汽车学习协调充电策略,以实现设定的多个目标。在调度过程中,每层网络的所有电动汽车向公共的通信信道发送本地状态信息,同时接收来自其他电动汽车的平均状态信息,即所有电动汽车通过访问此广播通信结构共享状态信息,然后将处理后的信息作为下一层网络的输入。

CommNet 的具体结构如图 1 所示。首先,输入所有电动汽车的状态,通过  $T^1$  和  $T^2$  两个通信步骤映射到各自的行为。在每个通信步骤中,模块  $f_n^m$ ,  $m=0,1,2$  传播隐藏状态向量  $h_n^m$ ,并在公共信道上广播由  $h_n^m$  取平均值得到的通信向量  $c_n^m$ ,然后输出向量  $h_n^{m+1}$ 。这两个向量的具体计算如下:

$$c_n^{m+1} = \frac{1}{N-1} \sum_{n' \neq n} h_{n'}^m, m=0,1 \quad (3)$$

$$h_n^{m+1} = f_n^m(h_n^m, c_n^m) = \tanh(H^m h_n^m + C^m c_n^m), m=0,1 \quad (4)$$

其中  $H^m$  和  $C^m$  为对应的系数矩阵。将上述公式合并,可以得到:

$$h_n^{m+1} = \tanh(T^m h_n^m) \quad (5)$$

$$\text{其中: } h_n^m = [h_1^m, \dots, h_N^m]^T, T^m = \begin{bmatrix} H^m & \dots & \tilde{C}^m \\ \vdots & \ddots & \vdots \\ \tilde{C}^m & \dots & H^m \end{bmatrix}, \tilde{C}^m = \frac{C^m}{N-1}。$$

图 1 右侧显示了  $f_n^m$  模块的内部细节。输出经过

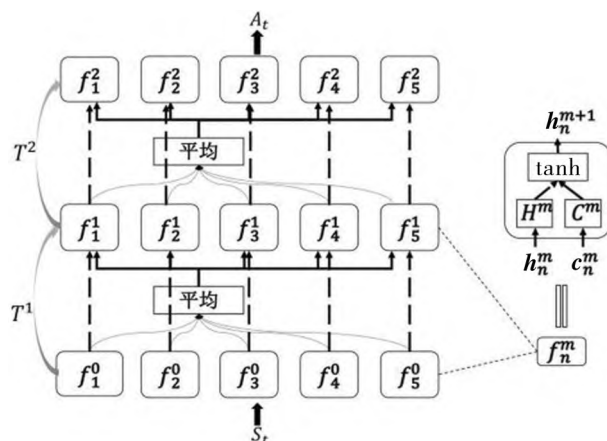


图 1 CommNet 结构

softmax 激活函数,得到最终的行为选择概率。根据该概率,在  $t$  时刻每个电动汽车在状态  $s_t^n$  下选择自己的行为  $a_t^n$ 。

## 3 深度强化学习框架

在机器学习和模式识别中,有许多经典的方法可用于函数逼近<sup>[10]</sup>。这里采用  $Q$  网络,它本质上是一个由三个稠密层组成的神经网络。 $Q$  网络的输入层完全连接到隐藏层,这层的输出值为:

$$Y_t = g(W_1 * X_t + B_1) \quad (6)$$

其中  $g$  为修正后的线性激活函数,  $W_1$ 、 $B_1$  分别为权重矩阵和偏置向量。类似地,隐藏层完全连接到输出层,输出值为最终近似到的行为值函数:

$$Q_w(S, A) = g(W_2 * Y_t + B_2) \quad (7)$$

其中  $W_2$ 、 $B_2$  分别为权重矩阵和偏置向量。

本文利用 critic 网络估计行为值函数:

$$Q^{\pi_a}(S, A) \approx Q_w(S, A) \quad (8)$$

同时,actor 网络根据 critic 部分获得的值指导策略函数的参数更新。整个算法的网络结构如图 2 所示。

值得注意的是,本算法在迭代过程中运用了两

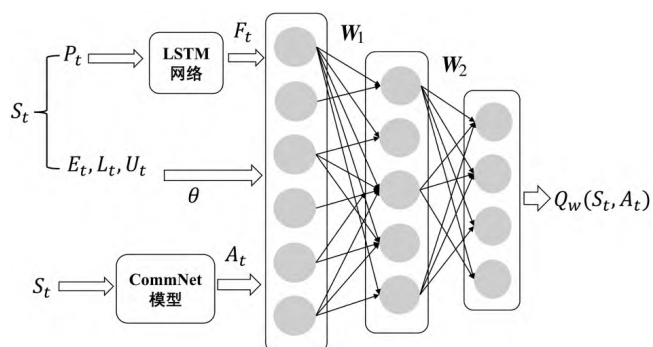


图 2 完整的算法网络图

个关键技术,一个是用于提高算法稳定性的目标网络,即引入目标 actor 网络和目标 critic 网络来固定参数,并生成与标签数据等价的目标值。另一个关键技术是经验回放,以存储采样的方式打破数据之间的相关性,避免连续一段时间内同一方向的梯度下降,从而保证算法的收敛性<sup>[11-12]</sup>。

完整的基于 CommNet 的深度强化学习算法的伪代码如下:

输入:折扣因子,学习率,迭代次数,更新周期;  
输出:行为。

1:初始化 actor 网络参数  $\theta$ 、critic 网络参数  $w$

2:初始化目标网络参数  $\hat{\theta}, \hat{w}$

3:for episode from 1 to  $M$  do

4:初始化系统环境

5:for time from 1 to  $T$  do

6:对于每个状态  $s_t^n \in S_t$ ,根据策略  $\pi_\theta(S_t, A_t)$  得出的概率  $\mu(S_t|\theta)$  选择行为  $a_t^n \in A_t$

7:在模拟环境中执行行为,然后反馈奖励值  $R_t$  和下一时刻的系统状态  $S_t'$

8:将转换对样本  $(S_t, A_t, R_t, S_t')$  存入经验回放池  $D$

9:If the number of samples  $< |D|$ , do

10:从  $D$  中随机抽样小批量的转换对  $F=\{(S_t, A_t, R_t, S_t')\}$

11:计算行为值函数的目标值:

$$\delta_t = R_t + \gamma Q_w(S_t', \mu(S_t'|\hat{\theta})) - Q_w(S_t, A_t)$$

12:根据随机梯度下降方法最小化 critic 网络的损失函数,以更新  $w$ :

$$\min L(w) = \frac{1}{\varphi} \sum_i [\delta_i - Q_w(S_i, A_i)]^2$$

13:利用策略梯度定理计算 actor 网络的梯度以更新  $\theta$ :

$$\nabla_\theta J(\theta) \approx \frac{1}{\varphi} \sum_i \nabla_\theta \mu(S_i|\theta) \nabla_\mu Q_w(S_i, \mu(S_i|\theta))$$

14:更新目标网络的参数:

$$\hat{w} = \eta w + (1 - \eta) \hat{w}$$

$$\hat{\theta} = \eta \theta + (1 - \eta) \hat{\theta}$$

15: end if

16: end for

17: end for

## 4 试验验证

本节给出了仿真结果,以验证所提出的基于 CommNet 的深度强化学习算法的性能。代码基于 Python 和 TensorFlow。

### 4.1 仿真环境设置

通过调查上海市的私家电动汽车用户在不同时刻的出行情况,模拟状态向量中的电动汽车位置变量。每辆电动汽车每天平均行驶 40 km。

假设研究的每个家庭都有一辆电动汽车。电动汽车的电池容量为 30 kW·h,平均能源消耗量为 0.15 kW·h/km,电动汽车平均充电率为 10 kW·h/h<sup>[13-14]</sup>。

在实现所提出的基于 CommNet 的深度强化学习算法时,为了稳定学习阶段,使用 Xavier 初始化器初始化由全连接层组成的 CommNet 的权值。另外,隐藏层节点数和通信信道节点数均设置为 128。在 critic 网络中,输入单元数等于状态向量、特征向量与行为向量的维数之和,输出单元数等于电动汽车的总数量。在训练过程中,折扣因子设置为 0.99, critic 网络和 actor 网络的学习率设置为 0.1,批量采样对的数目设为 8,迭代次数设为 2 000。在 CPU 为 i3-8100 的计算机上,训练过程大约在 1 250 次迭代后收敛,训练时间为 10 s 左右。

### 4.2 性能评估

#### 4.2.1 多目标协调充电

首先评估包含 3 辆电动汽车的小规模多智能体系统的性能。在某住宅的局部电网覆盖下,变压器为附近的 3 户家庭提供 30 kW·h 的电力。假设仿真的每个阶段都在一天内进行,时间间隔为 1 h。

图 3 显示了总奖励值随迭代次数的收敛情况。从中可看出奖励值在 1 500 次迭代左右开始收敛。下面验证本文提出的方法分别针对用户和电网的相关目标的实现效果。

一方面,对于用户,图 4(a)~图 4(c)依次显示了在算法达到收敛后的某一天内,每辆电动汽车观察到的实时电价和电动汽车位置,以及各自的行为决策曲线。从图中可以看出,如果电动汽车到达目的地时正处于谷时电价时段,且有足够的停车时长使荷电状态充至较饱和状态,则用户的充电意愿较强烈,如图 4(b)中 1~4 时、图 4(c)中 3~5 时显示的情况;反之,若电动汽车在区域的停驻时间较短,且正处于峰时电价时段,考虑到用户充电成本和停车时长的敏感程度,此时用户的充电概率相对较低,如



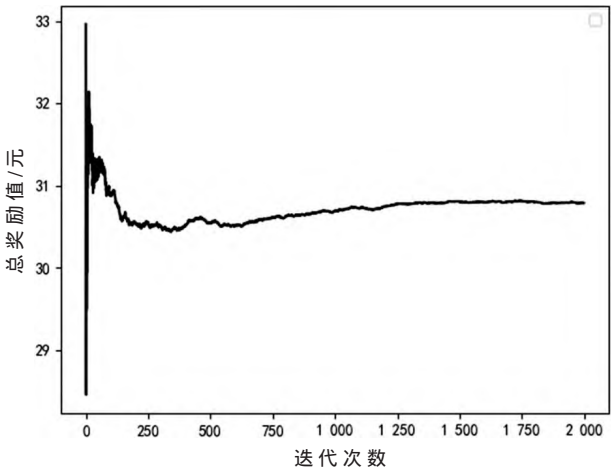


图3 所有电动汽车的总奖励值

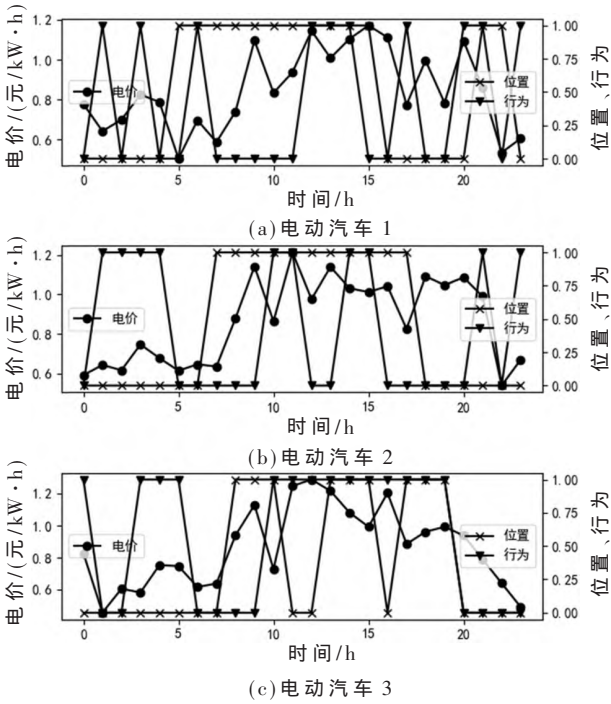


图4 各电动汽车的位置、电价与充电行为

图4(a)中14~16时、图4(b)中16~18时的情况。

另一方面,对于电网,图5(a)~5(c)依次显示了在算法达到收敛后的某一天内,每辆电动汽车接收到的变压器负载以及各自的行为决策曲线。如果变压器离满载还有足够的空间,则用户的充电意愿较强烈,如图5(a)中11~14时、图5(c)中13~18时的情况;反之,若变压器负载较高,那么此时用户的充电概率相对较低,如图5(b)中4~7时、图5(c)中19~21时显示的情况。

为了表明所提方法的优势,将其与Q-learning进行对比,结果如图6所示。可以看出,后者的总奖

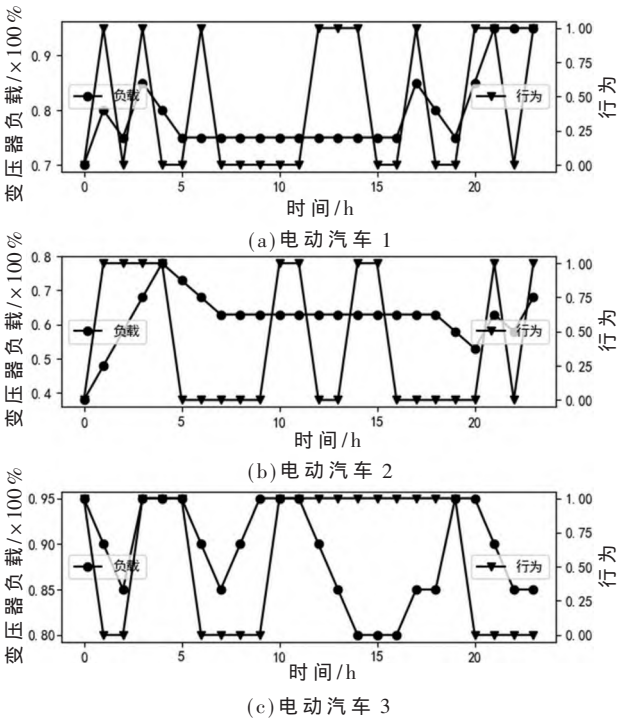


图5 各电动汽车的变压器负载与充电行为

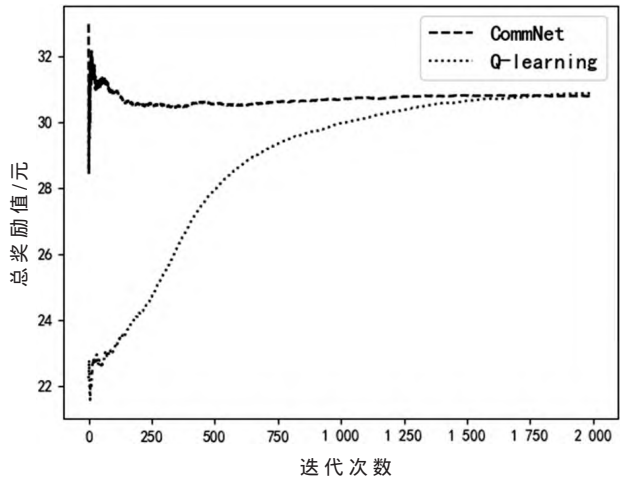


图6 CommNet 和 Q-learning 的奖励值收敛对比

励值收敛速度比CommNet方法慢很多,而且整体值也偏低。

通过以上讨论,可以充分说明本文提出的基于DRL的分布式算法是解决带有多个目标的协调电动汽车充电问题的有效方法。

#### 4.2.2 系统可扩展性

为了验证所提方法的可扩展性,将电动汽车的数量从3辆线性扩展到30辆。图7显示了随着电动汽车数量的增加,算法收敛所需的迭代次数。可

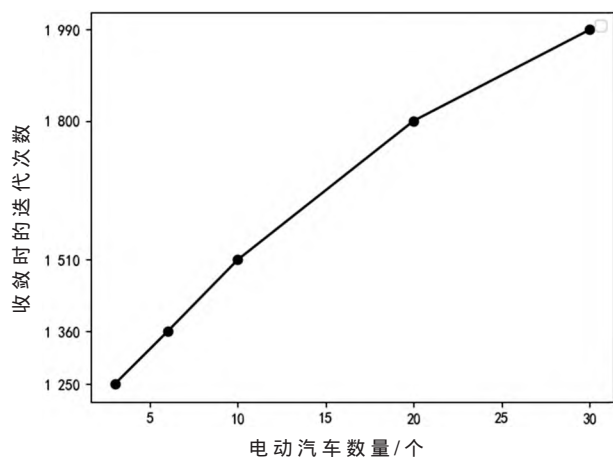


图7 电动汽车数量与算法收敛所需的迭代次数

以明显看出迭代次数呈现线性增长而非指数增长趋势,说明了该算法在处理大规模多智能体系统时的有效性。

接下来将基于 CommNet 的分布式方法与以下两种集中式基准方法做比较,进一步说明将 CommNet 网络作为策略方案的性能优势。

**随机策略:**在该模型中,每辆电动汽车有 50% 的概率不考虑系统状态,即其他电动汽车的状态,进行充电动作。

**贪婪策略:**它是最具代表性的一种传统策略,每个状态有  $\varepsilon$  的概率进行探索(即随机选取充或不充),而剩下的  $1-\varepsilon$  的概率则进行开发(选取当前状态下效用值较大的那个动作)<sup>[15]</sup>。

对于上述三种算法,在相同的驾驶模式下,设置相同的目标,模拟了 2 000 次的电动汽车协调充电训练来定义平均性能。为此,评估了在 3 辆、10 辆和 30 辆电动汽车的情况下,不同方法对奖励值收敛情况的影响,结果分别如图 8(a)~8(c)所示。可以看出,在三条曲线中,贪婪策略对应的曲线在训练初期波动较大,最终发散,未能收敛。而随机策略的训练曲线缓慢衰减,且对应的奖励值远小于 CommNet 方法下的奖励值。以上结果表明,基于 CommNet 的算法在收敛稳定性方面比其他两种集中式基准方法更加突出。此外,随着电动汽车数目的不断增多,所提算法的奖励值能在线性迭代次数内达到收敛。

综上所述,本文提出的基于 CommNet 的深度强化学习方法最终学习到了能达到预定多目标的充电策略,并取得了更好的性能。因此,它能有效地应用于多电动汽车协调充电系统,对电动汽车的动态

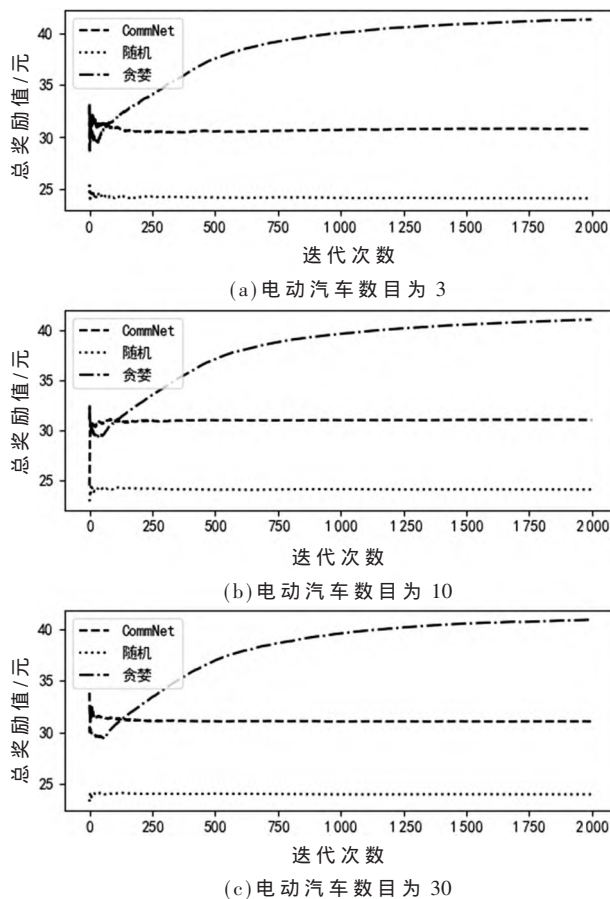


图8 不同数量电动汽车下三种策略的奖励值收敛情况

充电数据进行实时处理。

## 5 结论

本文针对包含多目标的多电动汽车协调充电问题,提出了一种分布式的多智能体深度强化学习方法。该方法的主要创新之处在于:首先,分别从用户和电网角度考虑,同时实现多个目标;其次,调度策略采用分布式方法,利用多辆电动汽车的局部信息,借助策略网络进行分布式计算以达成全局目标。通过数据密集仿真的性能评估,验证了所提算法的有效性和可扩展性。在未来,这项工作可以沿着不同的路线进一步扩展。

## 参考文献

- [1] XIANG Y, LIU J, LI R, et al. Economic planning of electric vehicle charging stations considering traffic constraints and load profile templates[J]. Applied Energy, 2016, 178: 647-659.
- [2] WEN C K, CHEN J C, TENG J H, et al. Decentralized plug-in electric vehicle charging selection algorithm in power systems[J]. IEEE Transactions on Smart Grid,

- 2012, 3(4): 1779–1789.
- [3] BAHRAMI S, PARNIANI M. Game theoretic based charging strategy for plug-in hybrid electric vehicles[J]. IEEE Transactions on Smart Grid, 2014, 5(5): 2368–2375.
- [4] SILVA F L D, NISHIDA C E H, ROIJERS D M, et al. Coordination of electric vehicle charging through multi-agent reinforcement learning[J]. IEEE Transactions on Smart Grid, 2020, 11(3): 2347–2356.
- [5] RODRIGUES T K, SUTO K, NISHIYAMA H, et al. Machine learning meets computation and communication control in evolving edge and cloud: challenges and future perspective[J]. IEEE Communications Surveys & Tutorials, 2020, 22(1): 38–67.
- [6] 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.
- [7] 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.
- [8] VAYA M G, ROSELLO L B, ANDERSSON G. Optimal bidding of plug-in electric vehicles in a market-based control setup[C]//Power Systems Computation Conference. IEEE, 2014.
- [9] SUTTON R, BARTO A. Reinforcement learning: an introduction[M]. Cambridge: MIT Press, 1998.
- [10] GRE F K, SRIVASTAVA R K, KOUTNIK J, et al. LSTM: a search space odyssey[J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, 28(10): 2222–2232.
- [11] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning[J]. Computer Science. arXiv: 1312.5602, 2013.
- [12] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518: 529–533.
- [13] 张美霞, 蔡雅慧, 杨秀, 等. 考虑用户充电差异性的家用电动汽车充电需求分布分析方法[J]. 电力自动化设备, 2020, 40(2): 154–163.
- [14] 刘坚, 金亨美. 电动汽车参与上海市电力需求响应潜力及经济性研究[J]. 供用电, 2017, 34(3): 26–31, 41.
- [15] 李琛, 李茂军, 杜佳佳. 一种强化学习行动策略  $\epsilon$ -greedy 的改进方法[J]. 计算技术与自动化, 2019, 38(2): 141–145.

(收稿日期: 2021-12-22)

## 作者简介:

张子霖(1997-), 通信作者, 女, 硕士研究生, 主要研究方向: 多智能体系统、深度强化学习。E-mail: zzzl@mail.ustc.edu.cn。

(上接第 64 页)

- [10] 章思宇, 邹福泰, 王鲁华, 等. 基于 DNS 的隐蔽通道流量检测[J]. 通信学报, 2013, 34(5): 143–151.
- [11] BUCZAK A L, HANKE P A, CANCRO G J, et al. Detection of tunnels in PCAP data by random forests[C]//Proceedings of the 11th Annual Cyber and Information Security Research Conference, 2016: 1–4.
- [12] NADLER A, AMINOV A, SHABTAI A. Detection of malicious and low throughput data exfiltration over the

DNS protocol[J]. Computers & Security, 2019, 80: 36–53.

- [13] SHERIDAN S, KEANE A. Detection of DNS based covert channels[C]//European Conference on Information Warfare and Security. Academic Conferences International Limited, 2015: 267.

(收稿日期: 2022-01-13)

## 作者简介:

陈解元(1990-), 男, 硕士, 工程师, 主要研究方向: 网络空间安全、安全态势感知。