doi:10.3772/j.issn.1002-0470.2024.07.009

基于双深度 Q 网络算法的多用户端对端能源共享机制研究①

武东昊②* 王国烽* 毛 磊** 陈玉萍** 张有兵*

(*浙江工业大学信息工程学院 杭州 310023)

(**浙江华云电力工程设计咨询有限公司 杭州 310026)

摘 要 端对端(P2P)电力交易作为用户侧能源市场的一种新的能源平衡和互动方式,可以有效促进用户群体内的能源共享,提高参与能源市场用户的经济效益。然而传统求解用户间 P2P 交易的方法依赖对于光伏、负荷数据的预测,难以实时响应用户间的源荷变动问题。为此,本文建立了一种以多类型用户为基础的多用户 P2P 能源社区交易模型,并引入基于双深度 Q 网络(DDQN)的强化学习(RL)算法对其进行求解。所提方法通过 DDQN 算法中的预测网络以及目标网络读取多用户 P2P 能源社区中的环境信息,训练后的神经网络可通过实时的光伏、负荷以及电价数据对当前社区内的多用户 P2P 交易问题进行求解。案例仿真结果表明,所提方法在促进社区内用户间 P2P 能源交易共享的同时,保证了多用户 P2P 能源社区的经济性。

关键词 端对端(P2P)能源共享;强化学习(RL);能源交易市场;双深度 Q 网络(DDQN)算法

近年来,随着可再生能源、储能系统等分布式能源的渗透率不断提高以及各类能源基础设施的不断建设,同时具有生产、消费电能特性的产消者用户(prosumer)在电力网络中逐渐兴起[1-4]。对于该类用户侧电力资源,传统的做法是按照法律规定的新能源优先采购政策,以固定价格或溢价采购的方式进行调度。但在大型储能设备建设成本较高、可再生能源补贴逐步降低的背景下,这种方式难以发展。而端对端(peer-to-peer,P2P)电力交易作为用户端能源系统中一种新的能量平衡交互方式,可以充分发挥产消者对于电力市场的调节作用,通过电能的商品属性,促进用户间的电力共享,降低用电成本,同时提升用户对可再生资源的就地消纳率[5-6]。

用户间的 P2P 交易问题,影响因素众多,各用户既会相互影响,又会不断观察和学习来调整自身行为,进而推动整个系统交易演化,整体是一个复杂适应性问题[7-9]。针对该问题,文献[10]考虑连续拍卖市场的价格波动,提出了一种自适应激进交易策略,通过连续双向拍卖机制完成 P2P 交易匹配。文献[11]考虑到现有多区域共享机制的不足及产消者在多区域能源共享中的复杂行为,提出了一种基于双层演化博弈的多区域点对点能源共享机制。文献[12]提出了一种基于双拍卖市场的分布式 P2P 电力交易方法,在不牺牲交易市场鲁棒性的前提下,实现了能源的协调互补。文献[13]以含光伏出力、储能系统和可控负荷的高渗透率产消者组成的配网为研究对象,利用交替方向乘子法进行分布式求解,通过有限的信息交互实现产消者间的 P2P 交易。上述研究设计的 P2P 能源共享机制,都由专门设立的第三方负责交易策略的制定和管理。但家庭等小规模用户之间的 P2P 电力交易,往往具有规模小、频率高、交易量变化大、参与者缺乏参与竞价过程的

① 国家自然科学基金(U22B20116)资助项目。

② 男,1997 年生,硕士生;研究方向:区域综合能源系统优化调度,分布式能源交易;联系人,E-mail:1942413736@qq.com。

(收稿日期:2022-12-13)

时间和意愿等特性。面对这一情况,第三方决策者即使能提出合理的交易策略也难以进行实时响应。

强化学习(reinforcement learning,RL)作为一种新型人工智能算法,通过学习策略来促进智能体与环境之间的交互,以最大化回报或实现特定目标[14]。其中智能体通过不断“试错”与环境交互,获得奖励来引导自我行为更新,通过累积奖励最大化获得最优控制策略。它是一种无模型的控制方案,在应用时不需要系统模型和对系统的先验知识。借助 RL 方法,可以在不进行优化计算和对市场模型了解不足的情况下进行市场交易行为,更适合小规模背景下的 P2P 能源交易系统。

目前,RL 在电力系统中的应用涵盖多个方面,包括需求响应管理、运行控制、经济调度等。例如文献[15]引入多智能体 Nash-Q 强化学习算法,将市场参与主体构建成智能体,经由智能体在动态市场环境中反复探索与试错寻找博弈均衡点。文献[16]提出了一种无模型的储能电池自优化经济运行系统,利用深度 Q 网络(deep Q network,DQN)求解电力系统最优充放电策略,获得最大周期增益。文献[17]提出了一种新的基于重复博弈的 P2P 能源交易框架,建立了平均效用最大化与最佳策略之间的关系。文献[18]在离散和连续的动作空间中,通过捕获负载需求和光伏剩余功率等信息,实现智能电网的最优控制。以上研究大多集中在电力市场中某一用户自身的行为,而未考虑其他参与交易用户的本地运行状态,不能适应去中心化多用户 P2P 交易的特点,也没有系统研究 P2P 电力交易背景下的用户交易决策问题。

为此,本文在已有研究的基础上以小规模家庭用户的 P2P 电力交易问题为研究对象,分析了 RL 技术在 P2P 电力交易应用中的关键问题,如交易参与者和电力交易价格模型、马尔科夫决策模型(Markov decision process,MDP)以及双深度 Q 网络(double deep Q network,DDQN)求解算法。最后通过案例分析,说明本文所提方法的有效性。本文的具体贡献如下。

(1)提出了一种适应小规模交易参与者的具有多类型用户的 P2P 能源社区交易框架。根据经济学原理来确定可以实时更新的用户间交易价格,更适合小规模 P2P 市场的特点。

(2)将多用户 P2P 交易问题等效为 MDP 模型,并针对问题构建了 MDP 中的状态空间、动作空间、奖励函数等元素。

(3)针对 MDP 模型中的储能设备动作和交易策略选择问题,采用 DDQN 算法进行分析求解。

(4)通过多用户 P2P 能源共享案例仿真,将所提算法与其他算法进行结果比较,验证了所提方法的有效性。

1　多用户 P2P 能源社区交易框架

本文根据多用户 P2P 能源共享社区内参与能源交易的用户特点,构建了如图 1 所示的多用户 P2P 能源交易框架示例。按照用户端设备的特点,将参与用户分为 3 类。

(1)消费者:只具有用户负荷的传统用户类型,不进行电力的产出与售卖,在共享结构中作为单纯的购买方存在。

(2)产消者 1 类:具有用户负荷以及光伏的用户类型,该类型用户的光伏产能优先选择内部消耗,其剩余部分参与社区内的电能交易。

(3)产消者 2 类:同时具有用户负荷、光伏以及储能设备的用户类型,该类型用户在满足自身用户负荷的基础上,可根据其他用户的电能需求/供应量,实时更新自身的购/售电策略。

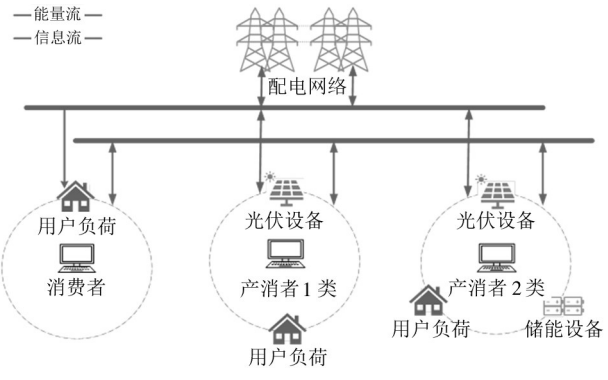


图 1　多类型用户 P2P 能源共享框架

1.1　各类型用户供需模型

对于多用户 P2P 能源共享社区内各类型用户,其具有如下的供需关系:

$$p^i_{\text{load}}(t) = p^i_{\text{op}}(t) + p^i_{\text{grid}}(t) + p^i_{\text{pv}}(t) + p^i_{\text{bes}}(t) \tag{1}$$

式中,$p^i_{\text{load}}(t)$ 为 t 时段内用户 i 需求负荷功率;$p^i_{\text{op}}(t)$ 为 t 时段内用户 i 与其他用户的交互功率,当 $p^i_{\text{op}}(t)$ 为正时表示向其他用户购买电能,当 $p^i_{\text{op}}(t)$ 为负时表示向其他用户售卖电能;$p^i_{\text{grid}}(t)$ 为 t 时段内用户 i 与电网的交互功率,当 $p^i_{\text{grid}}(t)$ 为正时表示向电网购买电能,当 $p^i_{\text{grid}}(t)$ 为负时表示向电网售卖电能;$p^i_{\text{pv}}(t)$ 为 t 时段内用户 i 的光伏输出功率;$p^i_{\text{bes}}(t)$ 为 t 时段内用户 i 储能设备的充放电功率,当 $p^i_{\text{bes}}(t)$ 为正时表示储能设备放电,当 $p^i_{\text{bes}}(t)$ 为负时表示储能设备充电。

消费者型用户因不具备光伏设备以及储能设备,其 $p^i_{\text{pv}}(t)$、$p^i_{\text{bes}}(t)$ 恒为 0;产消者 1 类用户因不具备储能设备,其 $p^i_{\text{bes}}(t)$ 恒为 0。

1.2 储能设备模型

产消者 2 类用户,可以通过储能设备采取不同的充放电策略,将盈余的光伏产出储存起来,在用电高峰时供自己使用以节约用电成本。此外,他们还可以通过灵活部署放电时间参与 P2P 交易,提高售电收入。不考虑温度和其他环境因素对储能设备充放电的影响,储能设备充放电的过程可表示为

$$E^i_{\text{soc}}(t) = E^i_{\text{soc}}(t-1) + \frac{\eta_{\text{bes}} p^i_{\text{bes}}(t)\Delta t}{Q_{\text{bes}}} \tag{2}$$

式中,$E^i_{\text{soc}}(t)$ 为 t 时段内用户 i 的储能设备荷电状态;Q_{bes} 为储能设备的容量;η_{bes} 为储能电池充放电系数,其表达式为

$$\eta_{\text{bes}} = \begin{cases} \eta_{\text{dis}} & p^i_{\text{bes}}(t) \geqslant 0 \\ \dfrac{1}{\eta_{\text{ch}}} & p^i_{\text{bes}}(t) < 0 \end{cases} \tag{3}$$

式中,η_{dis}、η_{ch} 分别为储能电池的放电效率和充电效率。

对于储能设备,还需避免设备深度充放电带来的损害,因此储能设备的荷电状态需要被限定在一定范围内:

$$E^{i,\min}_{\text{soc}} \leqslant E^i_{\text{soc}}(t) \leqslant E^{i,\max}_{\text{soc}} \tag{4}$$

式中,$E^{\min}_{\text{soc}}$、$E^{\max}_{\text{soc}}$ 分别为用户 i 的储能设备荷电状态上、下限。

1.3 交易价格模型

在电网和社区内用户的电能交易中,配电网络会根据社区内用户总负荷变化情况对各时段制定不同的电价水平,鼓励用电客户合理安排用电时间。本文中将电网提供的电价表示为

$$\lambda_{\text{sell}} = \{\lambda^1_{\text{sell}}, \cdots, \lambda^H_{\text{sell}}\} \tag{5}$$

$$\lambda_{\text{buy}} = \{\lambda^1_{\text{buy}}, \cdots, \lambda^H_{\text{buy}}\} \tag{6}$$

式中,λ_{sell} 为电网的售电价;λ_{buy} 为电网的购电价;H 表示时间段。根据当前的余电上网政策,电网对于用户的购售电价在一天中不进行变化。

对于多类型用户间的交易价格机制,本文对文献[10]、[19]中基于供需比(supply and demand ratio,SDR)的价格机制进行了改进。通过式(7)计算多用户 P2P 能源共享社区内的 SDR。

$$SDR(t) = \begin{cases} \dfrac{T_{\text{pv}}(t) + T_{\text{bes}}(t)}{T_{\text{load}}(t)} & T_{\text{bes}}(t) \geqslant 0 \\ \dfrac{T_{\text{pv}}}{T_{\text{load}}(t) + T_{\text{bes}}(t)} & T_{\text{bes}}(t) < 0 \end{cases} \tag{7}$$

其中,

$$T_{\text{pv}}(t) = \sum_{i=1}^{N} p^i_{\text{pv}}(t) \tag{8}$$

$$T_{\text{load}}(t) = \sum_{i=1}^{N} p^i_{\text{load}}(t) \tag{9}$$

$$T_{\text{bes}}(t) = \sum_{i=1}^{N} p^i_{\text{bes}}(t) \tag{10}$$

式中,$SDR(t)$ 为 t 时段内多用户 P2P 能源共享社区内的供需比;N 为多用户 P2P 能源共享社区内的用户总数;$T_{\text{pv}}(t)$ 为 t 时段内多用户 P2P 能源共享社区内各类型用户光伏输出总功率;$T_{\text{load}}(t)$ 为 t 时段内多用户 P2P 能源共享社区内各类型用户总负荷需求;$T_{\text{bes}}(t)$ 为 t 时段内多用户 P2P 能源共享社区内各类型用户储能设备充放电总功率。

根据 SDR 与交易价格呈负相关的基本经济规律,可得多用户 P2P 能源共享社区内各用户间的交易价格为

$$r_{\text{sell}}(t) = \begin{cases} \dfrac{\lambda^t_{\text{buy}}(\lambda^t_{\text{sell}} + \lambda)}{SDR(t)(\lambda^t_{\text{buy}} - \lambda^t_{\text{sell}} - \lambda) + \lambda^t_{\text{sell}} + \lambda} & 0 \leqslant SDR(t) < 1 \\ \lambda^t_{\text{sell}} + \dfrac{\lambda}{SDR(t)} & SDR(t) > 1 \end{cases} \tag{11}$$

$$r_{\text{buy}}(t) = \begin{cases} SDR(t)r_{\text{buy}}(t) + \lambda_{\text{buy}}^{t}(1 - SDR(t)) \\ \qquad 0 \leqslant SDR(t) < 1 \\ \lambda_{\text{sell}}^{t} + \lambda \qquad SDR(t) > 1 \end{cases} \tag{12}$$

式中，$r_{\text{sell}}(t)$、$r_{\text{buy}}(t)$ 分别为 t 时段社区内用户间的售、购电价格；λ 为光伏补偿价格，当 SDR 较大时用以维持光伏产消者的利益和积极性，同时也利于 P2P 交易的长期发展。

此外，为保证社区内 P2P 交易的正常进行，对其交易价格具有如下约束：

$$r_{\text{sell}}(t) \geqslant \lambda_{\text{sell}}^{t} \tag{13}$$

$$r_{\text{buy}}(t) \leqslant \lambda_{\text{buy}}^{t} \tag{14}$$

$$\lambda_{\text{buy}}^{t} \geqslant \lambda_{\text{sell}}^{t} \tag{15}$$

$$r_{\text{buy}}(t) \geqslant r_{\text{sell}}(t) \tag{16}$$

由式(7)～(16)可以看出，社区内用户的储能设备充电、放电行为是影响 P2P 交易价格的最重要因素。因此，用户储能设备的充放电策略是其在能源交易中获得最大利益的关键。

2 多用户 P2P 能源交易的马尔科夫决策模型

在社区内的多用户 P2P 能源交易框架中，$p_{\text{pv}}^{i}(t)$ 和 $p_{\text{load}}^{i}(t)$ 是在交易时采集的实时数据；$E_{\text{soc}}^{i}(t)$ 与储能设备上一时段的荷电状态以及当前时段的充放电动作相关；$r_{\text{sell}}(t)$、$r_{\text{buy}}(t)$ 基于 SDR 的定义，与 $p_{\text{pv}}^{i}(t)$、$p_{\text{load}}^{i}(t)$ 以及储能设备的运行状态相关。因此，用户调整 P2P 交易策略以获得最佳收益的过程可以被描述为马尔可夫决策过程 MDP。本文构建的多用户 P2P 能源交易 MDP 模型包括状态空间 S、动作空间 A、奖励函数 R 等元素，下文对其进行具体描述。

2.1 状态空间

对于在 t 时段内参与 P2P 交易的用户 i 的状态空间作如下定义：

$$S_{i,t} = \{p_{\text{pv}}^{i}(t),\ p_{\text{load}}^{i}(t),\ E_{\text{soc}}^{i}(t),\ r_{\text{sell}}(t),\ r_{\text{buy}}(t)\} \tag{17}$$

式中，$S_{i,t}$ 为 t 时段内用户 i 的状态空间；$r_{\text{sell}}(t)$、$r_{\text{buy}}(t)$ 根据 t 时段多用户 P2P 能源共享社区内的 $SDR(t)$ 进行计算，并根据当前时段内用户采取的 P2P 交易策略进行下一时段的状态更新。

2.2 动作空间

对于在 t 时段内参与 P2P 交易的用户 i 的动作空间作如下定义：

$$A_{i,t} = \{p_{\text{bes}}^{i}(t) \mid g\} \tag{18}$$

式中，$A_{i,t}$ 为 t 时段内用户 i 的动作空间；g 为动作空间的离散化程度，g 值越大动作空间包含的动作越少，g 值越小动作空间可以描述的动作越多。

2.3 奖励函数

在 MDP 中奖励函数负责引导用户挖掘状态空间中的决策相关因素并经过提炼后用于动作空间中进行动作选取。故本文将用户 i 在 t 时段内的购售电费用定义为奖励函数，奖励函数设置如下：

$$r_{i,t} = \begin{cases} \lambda_{\text{buy}}^{t}p_{\text{grid}}^{i}(t) + r_{\text{buy}}(t)p_{\text{op}}^{i}(t) & p_{\text{grid}}^{i}(t) \geqslant 0, p_{\text{op}}^{i}(t) \geqslant 0 \\ \lambda_{\text{buy}}^{t}p_{\text{grid}}^{i}(t) + r_{\text{sell}}(t)p_{\text{op}}^{i}(t) & p_{\text{grid}}^{i}(t) \geqslant 0, p_{\text{op}}^{i}(t) < 0 \\ \lambda_{\text{sell}}^{t}p_{\text{grid}}^{i}(t) + r_{\text{buy}}(t)p_{\text{op}}^{i}(t) & p_{\text{grid}}^{i}(t) < 0, p_{\text{op}}^{i}(t) \geqslant 0 \\ \lambda_{\text{sell}}^{t}p_{\text{grid}}^{i}(t) + r_{\text{sell}}(t)p_{\text{op}}^{i}(t) & p_{\text{grid}}^{i}(t) < 0, p_{\text{op}}^{i}(t) < 0 \end{cases} \tag{19}$$

式中，$r_{i,t}$ 为 t 时段内用户 i 的奖励函数。

对于 MDP 中的奖励函数，不仅需要考虑 t 时段内的奖励，还需考虑到未来的奖励。因此定义用户 i 在一个交易周期内的奖励函数为

$$R_i = \sum_{t=1}^{T} r_{i,t}$$

式中，R_i 为 t 时段内用户 i 的总奖励函数。

3 多用户 P2P 能源交易问题求解方法

考虑到多用户 P2P 交易市场通常具有交易规模小、频率高、波动性大的特点，本文采用基于深度强化学习的 DDQN 算法对多用户 P2P 能源交易问题进行求解。

DDQN 算法作为深度强化学习方法的一种，在处理小规模离散空间的问题中表现良好。其算法的主架构为预测网络(predict network)、目标网络(target network)和经验池 3 部分。

在算法的主架构中，存在 2 个循环操作：第 1 个循环由参与 P2P 交易的用户历史数据驱动，以 1 d

的调度动作为一条经验轨迹,进行循环迭代,并对经验池进行更新;第 2 个循环由经验池中提取批量经验轨迹数据进行驱动,并根据经验轨迹数据更新神经网络权重参数。DDQN 算法通过以上 2 个循环操作,对预测网络和目标网络进行训练。最终训练得到的网络,可输出当前状态下能够确保用户收益的实时交易策略,其训练过程由后文进行具体描述。

3.1 预测网络及目标网络训练

DDQN 算法中的神经网络,其作用是对用户在状态 S 下采取动作 a 的价值,即动作值函数 $Q(S,a)$ 进行近似:

$$Q(S,a,\theta) \approx Q(S,a) \tag{19}$$

DDQN 算法通过 Q 学习算法获得神经网络可学习的目标函数,即构建神经网络可优化的损失函数:

$$L(\theta) = E[(\text{Target } Q - \text{Predict } Q)^2] \tag{20}$$

式中,θ 为神经网络的权重参数;Predict Q 为预测网络输出的预测 Q 值;Target Q 为目标神经网络输出的目标 Q 值。Predict Q 可表示为

$$\text{Predict } Q = Q(S_{i,t},\ a_{i,t},\ \theta_i) \tag{21}$$

式中,θ_i 表示神经网络内输入和输出之间的关系;$a_{i,t}$ 为用户 i 的预测神经网络根据 t 时段的状态 $S_{i,t}$,从动作空间 $A_{i,t}$ 中选取的储能设备充放电动作。当微能源系统执行动作 $a_{i,t}$ 后,获得奖励 $r_{i,t}$,同时系统进入下一时段的环境状态。Target Q 可表示为

$$\text{Target } Q = r_t + \gamma \max_{a_{i,t+1}} Q(S_{i,t+1},a_{i,t+1},\theta_i^-) \tag{22}$$

式中,γ 为未来的 Q 值在当前时刻的衰减率;$S_{i,t+1}$ 为用户 i 在 $t+1$ 时段的 P2P 交易状态;a_{t+1} 为用户 i 的目标神经网络根据 $t+1$ 时段的状态 $S_{i,t+1}$,从动作空间 A 中选取的使动作值函数 Q 最大的调度动作。

在获得损失函数后,采用 Adam 算法(adaptive moment estimation)对神经网络损失函数模型 $L(\theta)$ 的权重参数 θ 进行求解,并将更新后的权重参数 θ 复制给预测神经网络。经过固定轮次迭代后,将预测神经网络的相关参数复制给目标网络,保持一段时间内目标 Q 值不变,降低预测 Q 值和目标 Q 值的相关性,提高算法稳定性。

3.2 经验回放

DDQN 算法具有独特的经验池回放机制,在进行每一步循环操作时会将用户和 P2P 交易框架交互得到的样本信息即当前状态、当前选取动作、当前动作获得奖励、下一时刻状态及布尔值存储于经验池中,当需要对预测网络和目标网络训练时,从经验池中随机抽取小批量的历史经验样本数据来对神经网络参数进行训练。

每个经验样本以如下 $(S_{i,t},a_{i,t},r_{i,t},S_{i,t+1},done)$ 五元组的形式存储到经验池中,其中,$done$ 为布尔值类型,表示用户 i 在 $t+1$ 时段的状态 $S_{i,t+1}$ 是否为终止状态。P2P 交易框架中的用户每执行一步后,需要把执行该步所获得的经验信息存储于经验池。在执行数步后,从经验池中随机抽小批量经验样本数据,输入到预测网络和目标网络中。基于抽样的经验样本数据,执行式(22),对预测网络和目标网络中的参数 θ_i、θ_i^- 进行更新。本文构建的多用户 P2P 交易框架中各用户的预测网络、目标网络具体更新训练流程如图 2 所示。

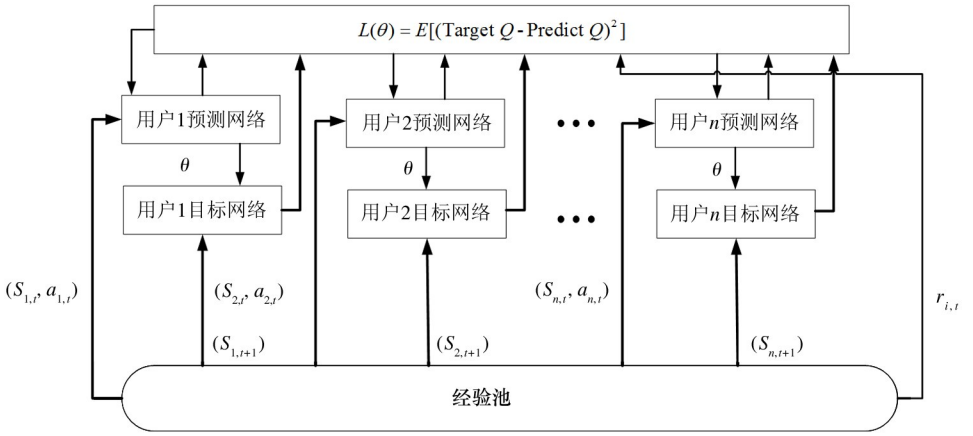


图 2 DDQN 神经网络训练过程

4　实验验证与分析

为了验证本文所提方法适用于多用户 P2P 能源共享,本实验基于 Python 实现算法编写,计算机配置为 CPU Intel Core i5、内存为 8 GB。

4.1　案例设置

采用包含 2 个消费者类用户,3 个产消者 1 类用户和 3 个产消者 2 类用户的多用户 P2P 能源共享社区作为研究对象。其中产消者 2 类用户,采用基于 DDQN 算法的储能设备行动策略,另 2 种用户因不具备储能设备终端采用线性最优互动策略。多用户 P2P 能源共享拓扑结构和本地设备配置如图 3 所示,交易时段为从 0:00—24:00 的 24 个时段。

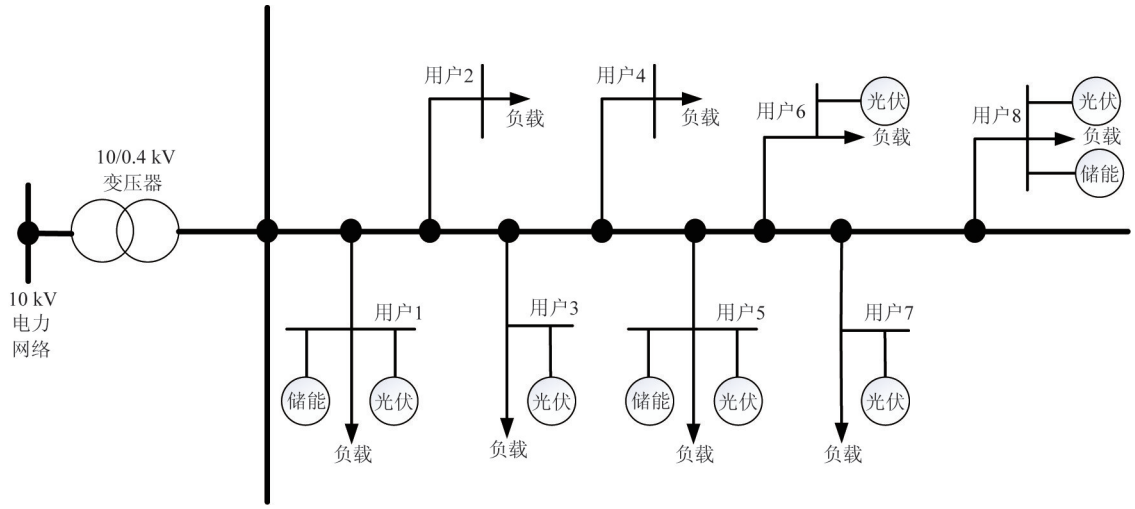


图 3　多用户 P2P 能源共享拓扑图

4.2　训练数据与神经网络参数设置

本文采用的 DDQN 算法在应用于多用户 P2P 能源交易前,需要先通过历史数据对其神经网络进行训练,以得到适配于交易环境的网络参数。训练采用的历史数据为某微能源网社区 1 月—12 月的实际光伏设备出力和负荷需求,其光伏平均出力及负荷平均功率数据如图 4 和图 5 所示。

以1月1日00:00时为起始,用户预测网络和目标网络接收来自 IES 环境的状态信息,然后根据第3节所述的学习过程进行循环迭代,更新神经网络参数,直至训练结束。训练时采用的电价数据如图 6所示。

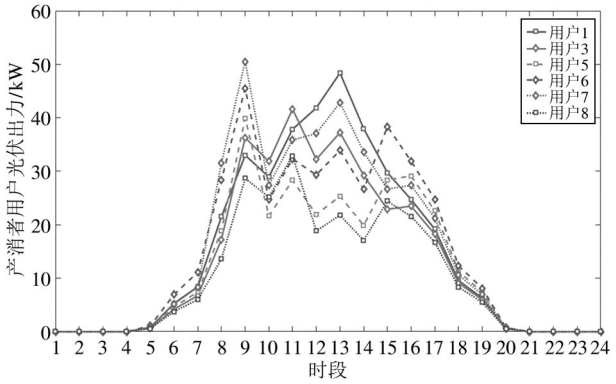


图 4　光伏出力样本数据

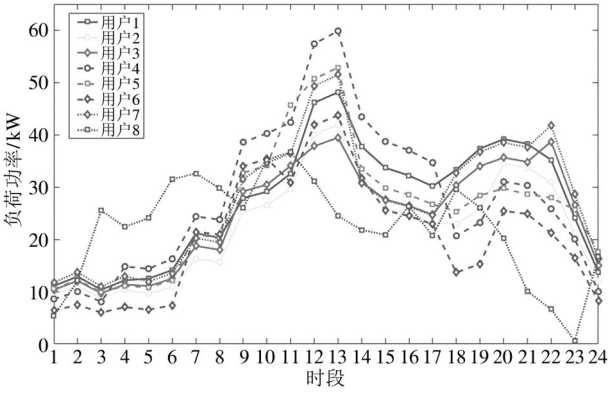


图 5　负荷需求样本数据

经过多次尝试,本文设定 DDQN 算法中的超参数如表 1 所示,经验池的样本存储量为 320 000,每次小批量采样规模为 32,初始探索率为 0.100,最终探索率为0.001,探索步数为1 600 000,学习率取

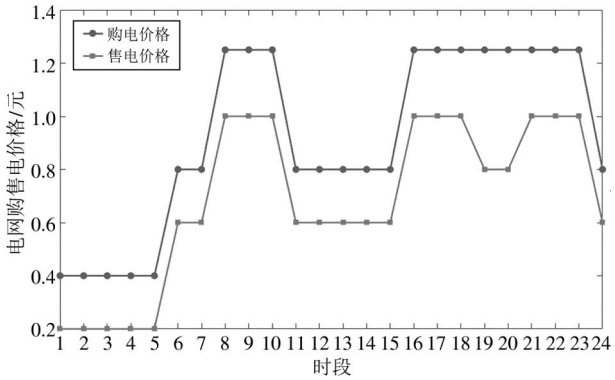


图 6　电网购售电价格

表 1　DDQN 算法相关超参数

超参数	数值
样本存储量	320 000
小批量样本	32
初始探索率	0.100
最终探索率	0.001
探索步数	1 600 000
学习率	0.010
动作离散度	54

0.01,动作离散度为 54,且每训练 10 次更新一次神经网络网络参数。

4.3　DDQN 神经网络训练及动作策略

本文所提产消者 2 类用户的预测网络和目标网络为结构相同的 3 层网络,各层神经元个数分别为 256、256、128。通过历史数据对神经网络参数进行迭代更新,当固定间隔达到 10 000 步数时,在训练数据外,采取 1 组数据为测试集,通过观察交易用户在测试集数据上进行实时调度的平均奖励,分析其是否已经学会合理、有效的调度策略。平均奖励的计算公式为

$$\bar{R} = \frac{\sum_{m=1}^{M} R}{M} \tag{23}$$

式中,m 表示开始调度天数;M 为调度天数;$\bar{R}$ 为在调度天数内各用户对测试集进行调度所获得的平均奖励总和。通过观察平均奖励的变化,可以对智能体的学习情况进行了解,其平均奖励变化过程如图 7所示。

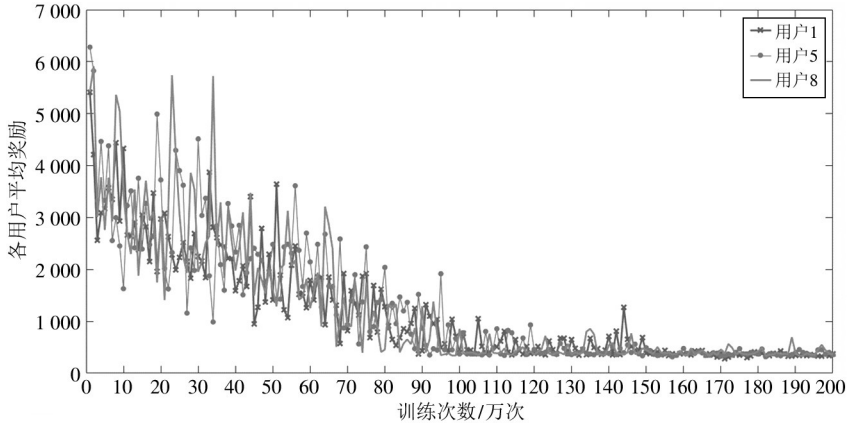


图 7　各用户平均奖励收敛曲线

从平均奖励变化曲线可以看出,各用户平均奖励在迭代约 1 500 000 次时趋于稳定,此时可以认为各用户已学会有效的调度策略。

在上述各用户神经网络训练完毕的基础上,采用训练样本外某一日内的数据进行多用户 P2P 能源共享,得到产消者 2 类用户在该日的动作策略结果即储能设备荷电状态变化,如图 8 所示。

储能设备不会采取过度充、放电的越限动作,且在一个调度周期内最终会回到与调度周期初相近的电池荷电状态,保证多用户 P2P 能源交易可持续稳定运行。

4.4　结果分析对比

为验证本文方法求解多用户 P2P 能源交易问题的优势,与传统 DQN 算法求解获得结果以及由重拟线性化技术(reformulation linearization technique, RLT)对模型进行线性化处理后,再通过 CPLEX 求

解获得的结果进行对比分析。不同方法下求解各用户交易结果如图 9 所示。

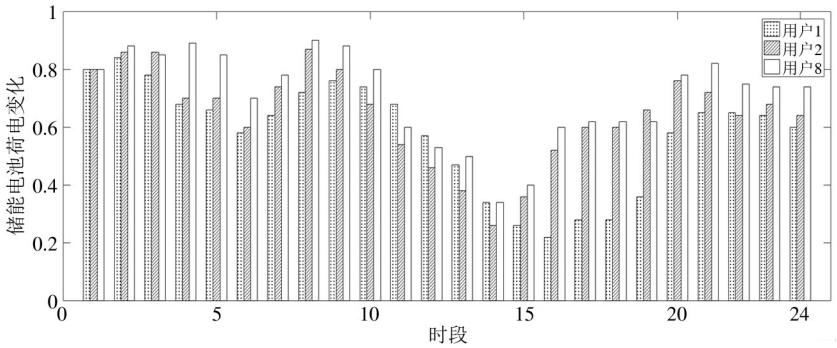
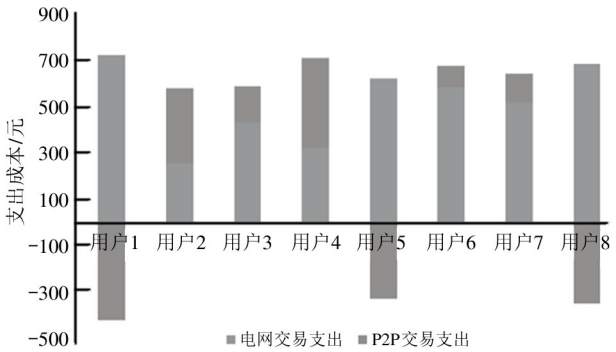
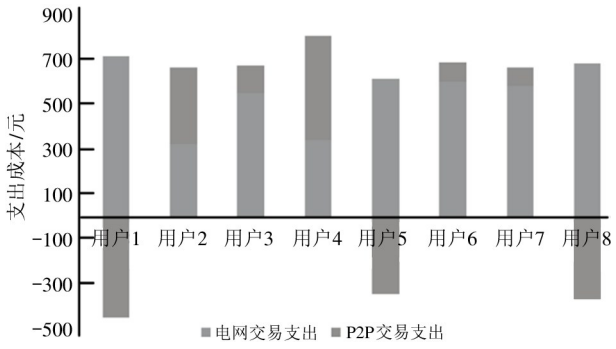


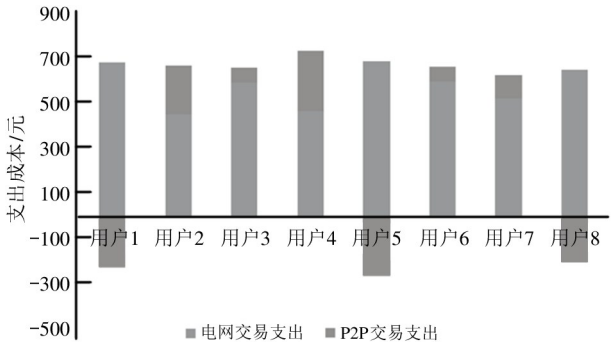
图 8　各用户储能设备的荷电状态



（a）基于 DDQN 算法的交易结果



（b）基于传统 DQN 算法的交易结果



（c）基于 RLT 重构的交易结果

图 9　各算法求解交易结果

各方法求解的多用户 P2P 能源共享社区购电总支出对比如表 2 所示。因为传统 DQN 算法相较于 DDQN 算法存在过估计问题，即估计的值函数比真实值函数要大。这使得传统 DQN 算法求解所得动作策略相较于 DDQN 算法，增加了对于不具备储能设备终端的用户购电支出。而经 RLT 处理后进行求解的方法相较于 DDQN 算法，无法顾及多用户 P2P 能源交易中的非线性关系，这使其不能充分调动社区内的储能设备、合理避开用电高峰以缩减社区购电成本。因此，相较其他 2 种方法，基于 DDQN 算法得到的结果更优，验证了所提方法在促进多用户 P2P 交易共享的同时，减少了社区购电的总支出，保证了多用户 P2P 能源社区的经济性。

表 2　不同算法下的多用户 P2P 能源社区购电总成本

方法	多用户 P2P 能源社区购电总成本/元
DDQN 算法	3 633
DQN 算法	3 909
RLT 重构	4 362

5　结 论

本文以含多用户类型的 P2P 能源交易社区为研究对象，提出一种基于 DDQN 算法的多用户 P2P 能源社区交易框架，在保证多用户 P2P 能源社区经济性的前提下，用于社区内各类型用户间的电能交

易。所提算法有效地避免了传统求解方式中依赖光伏、负荷预测无法实时求解随机变动的问题。通过仿真实验,将训练后的预测网络以及目标网络用于测试集数据进行求解,将其获得的交易结果以及社区购电总成本与其他算法进行比较,证明了算法的有效性。本文在实现多用户 P2P 能源共享时,仅考虑了用户侧的设备终端特性,而对配网侧的设备特性缺乏考虑。因此后续将继续研究如何构建合理、有效的园区交互市场。

参考文献

[1] 黎静华, 朱梦姝, 陆悦江, 等. 综合能源系统优化调度综述[J]. 电网技术, 2021,45(6):2256-2272.

[2] 文云峰, 杨伟峰, 汪荣华, 等. 构建 100% 可再生能源电力系统述评与展望[J]. 中国电机工程学报, 2020, 40(6):1843-1856.

[3] 安麒, 王剑晓, 武昭原, 等. 高比例可再生能源渗透下的电力市场价值分配机制设计[J]. 电力系统自动化, 2022,46(7):13-22.

[4] 肖云鹏, 王锡凡, 王秀丽, 等. 面向高比例可再生能源的电力市场研究综述[J]. 中国电机工程学报, 2018,38(3):663-674.

[5] 张虹, 闫贺, 申鑫, 等. 面向能源社区能量管理的配网产消者分布式优化调度[J]. 中国电机工程学报, 2022,42(12):4449-4459.

[6] 马丽, 刘念, 张建华, 等. 基于主从博弈策略的社区能源互联网分布式能量管理[J]. 电网技术, 2016,40(12):3655-3662.

[7] 唐成鹏, 张粒子, 刘方, 等. 基于多智能体强化学习的电力现货市场定价机制研究(一):不同定价机制下发电商报价双层优化模型[J]. 中国电机工程学报, 2021,41(2):536-553.

[8] 张粒子, 唐成鹏, 刘方, 等. 基于多智能体强化学习的电力现货市场定价机制研究(二):结合理论与仿真的定价机制决策框架[J]. 中国电机工程学报, 2021, 41(3):1004-1018.

[9] 赵天辉, 王建学, 陈洋. 面向综合能源交易的新型城镇分层市场架构和出清算法[J]. 电力系统自动化, 2021,45(4):73-80.

[10] LIU N, YU X, WANG C, et al. Energy-sharing model with price-based demand response for microgrids of peer-to-peer prosumers[J]. IEEE Transactions on Power Systems, 2017,32(5):3569-3583.

[11] 何鑫雨, 董萍, 刘明波, 等. 基于双层演化博弈模型的多区域点对点能源共享机制[J]. 电网技术, 2023, 47(1):163-174.

[12] WANG Z, YU X, MU Y, et al. A distributed peer-to-peer energy transaction method for diversified prosumers in urban community microgrid system[J]. Applied Energy, 2020,260:114327.

[13] 陈修鹏, 李庚银, 周明, 等. 考虑新能源不确定性和点对点交易的配网产消者分布式优化调度[J]. 电网技术, 2020,44(9):3331-3340.

[14] ZHANG Z, ZHANG D, QIU R C. Deep reinforcement learning for power system applications: an overview[J]. CSEE Journal of Power and Energy Systems, 2019,6(1):213-225.

[15] 孙庆凯, 王小君, 王怡, 等. 基于多智能体 Nash-Q 强化学习的综合能源市场交易优化决策[J]. 电力系统自动化, 2021,45(16):124-133.

[16] 冯昌森, 张瑜, 文福拴, 等. 基于深度期望 Q 网络算法的微电网能量管理策略[J]. 电力系统自动化, 2022,46(3):14-22.

[17] WANG H, HUANG T, LIAO X, et al. Reinforcement learning in energy trading game among smart microgrids [J]. IEEE Transactions on Industrial Electronics, 2016, 63(8):5109-5119.

[18] HIRATA T, MALLA D B, SAKAMOTO K, et al. Smart grid optimization by deep reinforcement learning over discrete and continuous action space[J]. Bulletin of Networking, Computing, Systems, and Software, 2019,8(1):19-22.

[19] LONG C, WU J, ZHOU Y, et al. Peer-to-peer energy sharing through a two-stage aggregated battery control in a community microgrid[J]. Applied Energy, 2018,226: 261-276.

Research on multi-user P2P energy sharing mechanism based on DDQN algorithm

WU Donghao[*], WANG Guofeng[*], MAO Cui[**], CHEN Yuping[**], ZHANG Youbing[*]

([*]School of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)

([**]Zhejiang Huayun Electric Power Engineering Design Consulting Co, Ltd, Hangzhou 310026)

Abstract

As a new way of energy balance and interaction in the user end energy market, peer-to-peer(P2P) power trading can effectively promote the energy sharing within the user group and improve the economic benefits of the users participating in the energy market. However, the traditional method of solving P2P power trading can not respond to the change of the source load among users in real time. Therefore, this paper establishes a multi-user P2P energy community trading model based on multi-type users, and introduces the deep reinforcement learning (RL) algorithm based on double deep Q network (DDQN) to solve it. The proposed method reads the environmental information in the multi-user P2P energy community through the prediction network and the target network in the DDQN algorithm. The trained neural network can solve the multi-user P2P trading problem in the current community through the real-time photovoltaic, load and electricity price data. Finally, the simulation results prove that the proposed method not only promotes the sharing of P2P energy trading among users in the community, but also ensures the economy of the multi-user P2P energy community.

Key words: peer-to-peer(P2P) energy sharing, reinforcement learning(RL), energy trading market, double deep Q network(DDQN)