

考虑用户禀赋效应的需求响应增量激励策略

华昊辰¹, 贾韵思¹, 陈星莺¹, 官家凯¹, 刘迪², 余昆¹, 甘磊¹

(1. 河海大学电气与动力工程学院, 江苏省 南京市 211100;
2. 清华大学电机工程与应用电子技术系, 北京市 海淀区 100084)

Incremental Incentive Strategy for Demand Response Considering User Endowment Effect

HUA Haochen¹, JIA Yunsi¹, CHEN Xingying¹, GONG Jiakai¹, LIU Di², YU Kun¹, GAN Lei¹

(1. School of Electrical and Power Engineering, Hohai University, Nanjing 211100, Jiangsu Province, China;
2. Department of Electrical Engineering, Tsinghua University, Haidian District, Beijing 100084, China)

ABSTRACT: Developing demand response strategies for users is one of the effective ways for electricity retailers to cope with price risks in the spot market. However, existing demand response strategies have not fully considered the differences in user response elasticity and psychological behavioral characteristics, resulting in an imbalance in consumer surplus allocation and limited incentive effects. Therefore, this article proposes a demand response incremental incentive strategy that considers the differences in user endowment effects and response elasticity. Firstly, by introducing endowment effect factors to quantify the subjective psychological cost of users' existing electricity rights, a time-varying incremental incentive function that dynamically matches the marginal cost and psychological loss of users is constructed to avoid excessive allocation of consumer surplus in the initial response stage. Secondly, to resolve the asynchronous coupling issue between the retailer's day-ahead bidding and intraday incentive strategies, this paper designs a novel asynchronous coupling solution method based on curriculum learning-enhanced twin delayed deep deterministic policy gradient (TD3) algorithm. This method progressively learns complex high-uncertainty environments—such as user behavior and market electricity prices—from simple to challenging stages, achieving coordinated optimization of asynchronous coupling problems. Theoretical analysis shows that the proposed incremental incentive strategy can reduce the unit incentive cost of electricity retailers and enhance the depth of high elasticity user response. Finally, the simulation results indicate that the proposed strategy dynamically redistributes consumer surplus

by following changes in user marginal costs, reducing excessive consumer surplus in the initial response phase and achieving fair allocation of incentive resources. In addition, the proposed algorithm has significantly improved convergence compared to the deep deterministic policy gradient (DDPG) algorithm, and its convergence speed has increased by 21.88% compared to the TD3 method without curriculum learning.

KEY WORDS: endowment effect; deep reinforcement learning; response elasticity; asynchronous coupling optimization; incremental incentive strategy

摘要: 针对用户制定需求响应策略是售电商应对现货市场价格风险的有效途径之一。然而, 现有需求响应策略未充分考虑用户响应弹性差异与心理行为特征, 导致消费者剩余分配失衡与激励效果受限。为此, 提出一种考虑用户禀赋效应与响应弹性差异的需求响应增量激励策略。首先, 通过引入禀赋效应因子量化用户对已有用电权的主观心理成本, 构建动态匹配用户边际成本与心理损失的时变增量激励函数, 避免响应初期消费者剩余过度分配。其次, 针对售电商日前投标与日内激励策略的异步耦合问题, 设计基于课程学习改进的孪生延迟深度确定性策略梯度(twin delayed deep deterministic policy gradient, TD3)的异步耦合求解方法, 由易到难地分阶段逐步学习用户行为、市场电价等高不确定性复杂环境, 实现异步耦合问题的协同优化。理论分析表明, 所提增量激励策略可降低售电商单位激励成本并提升高弹性用户响应深度。最后, 仿真结果表明, 所提策略通过跟随用户边际成本变化来动态再分配消费者剩余, 减少响应初期的过度消费者剩余, 实现了激励资源公平配置。此外, 所提算法相比深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法收敛性显著提高, 且较无课程学习的TD3方法收敛速度提升了21.88%。

关键词: 禀赋效应; 深度强化学习; 响应弹性; 异步耦合优化; 增量激励策略

基金项目: 国家重点研发计划项目(2022YFE0140600); 国家自然科学基金项目(52377093)。

National Key R&D Program of China (2022YFE0140600); Project Supported by National Natural Science Foundation of China (52377093).

0 引言

随着电力体制改革进一步发展,中国已经陆续开展现货电力市场交易试点工作。现货电力市场电价由市场上所有参与主体的供需平衡决定,导致现货市场电价波动较大^[1]。现货市场电价波动使得售电商需要承担相应的风险,进行现货市场风险管理对售电商尤为重要。面向用户实施需求响应策略是售电商应对现货市场风险的一种重要手段^[2]。因此,售电商基于当前背景,如何合理制定投标策略满足用户用电需求及采用有效的需求响应手段来引导用户调整自身行为值得重点研究。

售电商在面临价格倒挂即购电电价高于售电价格的时段时能够通过需求响应减少自身的售电损失,包括基于价格的需求响应和基于激励的需求响应。基于价格的需求响应又包括实时电价^[3]与分时电价等形式^[4]。实时电价指售电商将电价风险转移给用户,这将导致用户参与需求响应的积极性不高。分时电价是常用的方法,但一段时间内电价不再变化,售电商仍是现货市场价格风险的主要承担者。如果仅采用基于价格的需求响应,售电商将陷入规避现货市场风险与吸引用户购电的两难境地^[5]。为此,本文在用户广泛接受的分时电价的基础上,考虑一种激励策略来有效引导用户积极地参与需求响应。

激励型需求响应研究中,售电商通过激励补贴引导用户改变负荷需求,当激励补贴能够匹配用户因响应产生的成本时,用户将参与响应^[6]。用户参与需求响应时,其响应弹性各不相同,应充分考虑用户响应弹性的差异。目前,大部分研究都是根据消费者特点分类对不同类型消费者展开需求响应研究。文献[7]以电动汽车聚合商为研究对象建立激励折扣模型。有些学者针对居民用户提出了一种价格信号与积分制联合激励下的需求响应优化策略^[8-9]。文献[10]针对工商业可调负荷资源,考虑响应激励补贴机制来聚合可调节资源。上述研究虽然考虑了电力用户差异化响应特性,然而,采用的激励策略总是与用户响应总量耦合。根据微观经济学中用户行为理论,用户由于响应产生的边际成本随着响应量的增加而增加。为此,激励补偿单价取决于目标响应总量下的最高边际成本,无法跟踪用户边际成本的变化,导致响应初期过度的消费者剩余,即消费者愿意支付的最高价格与实际支付的市

场价格之间的差额。

在此基础上,目前有学者基于用户需求响应量,提出包含区间长度、补偿基价和价格增长率 3 个参数的阶梯型需求响应激励机制^[11-12]。虽然阶梯型需求响应激励机制通过响应量分段阈值实现补偿价格的阶梯式提升,然而,这既难以适应电力市场实时电价的动态波动,也无法精准匹配用户响应弹性等差异化特性。因此,亟需提出一种考虑用户响应弹性的激励措施,能跟随用户边际成本变化,有效避免初期消费者剩余过度并实现用户间激励再分配。

现有激励策略研究多聚焦于激励价格与形式的优化,却鲜有关注用户响应意愿对策略有效性的影响。传统研究基于微观经济学理论,默认用户为完全理性决策者,而行为经济学研究表明用户实际决策具有显著有限理性特征。因此,售电商有必要将用户有限理性行为与动态响应特性纳入激励策略设计。目前,已有学者考虑将用户心理因素纳入需求响应模型中,但未考虑其对增量激励策略的影响^[13]。有学者提出了考虑有限理性用户选择行为的定制化电价套餐设计方法^[14]。文献[15]分析了用户在激励型需求响应中的有限理性行为。现有关于用户响应行为的文献大都研究负荷反弹效应^[16-17]、用户动态响应特性^[18]及用户响应行为不确定^[19-21]。上述研究尚未考虑有限理性行为对增量激励策略的作用机制,导致激励资源分配与用户心理成本脱节。因此,缺少一种计及有限理性用户响应行为特性影响的激励策略。

另一方面,售电商需在日前电力批发市场投标向用户提供电力供应服务^[22],其日前投标策略的制定受日内需求响应策略的影响,而后者在日前阶段属于未知信息。由于 2 类策略存在时空异步耦合关系即日前投标与日内激励分属不同时段,有必要提出有效解决信息不完全的优化策略。研究提出了多种优化方法,主要分为 2 类传统方法:基于数学规划的确定性方法^[23]和基于规则的启发式搜索算法^[24-25]。前者通过构建精确模型实现全局最优,但对动态扰动适应性差;后者虽能处理非线性问题且不依赖严格的数学模型,而是通过人工设计的搜索规则进行优化,但在多时间尺度优化中易陷入局部最优。考虑到用户的多样性,传统的方法难以建立合适的模型并确定相应的参数来准确描述每个用户的响应行为。相比之下,数据驱动的方法通

过训练历史数据得出最优决策,而不需要显式的模型^[26]。可采用无模型深度强化学习算法(deep reinforcement learning, DRL)解决用户精细化响应模型缺失环境下的时序决策优化问题。

针对上述问题,本文提出一种可动态跟踪考虑用户禀赋效应心理因素后的边际响应成本变化的增量激励策略,制定非线性时变单位激励价格,以提升响应效率。具体地,首先,介绍市场机制及研究框架,厘清售电商与用户互动关系;进而,提出增量激励策略并建立售电商日前投标及日内售电模型、考虑用户心理和动态响应特性的需求响应模型;接着,解析考虑用户响应弹性和禀赋效应的增量激励策略的动态匹配机制,并理论层面论证策略优越性;最后,针对电价预测误差与用户行为不确定性等多源干扰,提出基于课程学习改进的孪生延迟深度确定性策略梯度(twin delayed deep deterministic policy gradient, TD3)算法,通过分阶段的学习过程,使智能体能够逐步适应复杂的不确定性环境,以解决日前投标与日内激励策略的异步耦合优化问题,最后,通过仿真验证其有效性。本文主要贡献如下:

1) 为提升售电商在激励策略制定过程中的有效性,本文建立一种综合考虑用户心理因素和动态响应特性的需求响应模型。基于行为经济学的理论框架,本文从心理学视角解释用户的有限理性行为,为构建用户行为模型提供新视角。具体地,采用贝叶斯转移概率刻画负荷反弹效应中的不确定性,以描述用户当前时段的响应策略对未来时段响应策略的动态影响。同时,引入禀赋效应因子来量化用户因禀赋效应产生的心理偏好,从而更准确地表征用户对用电权的主观价值评估及其对需求响应行为的潜在影响。

2) 面对售电商在价格倒挂时售电造成损失的问题,本文提出一种考虑用户禀赋效应与响应弹性差异的增量激励策略。通过数据驱动方法考虑到用户心理因素,为售电商提供有效的激励策略,可以合理地抵消用户参与需求响应时产生的心理需求,在提升用户响应参与度的同时增加售电商收益。此外,该策略解耦每单位激励成本的增量与用户总响应量,实现了激励补偿与用户响应弹性特性的动态匹配。

3) 提出一种基于课程学习改进的 TD3 算法的异步耦合优化策略。考虑到售电商在日前投标时受

到日内激励策略影响的用户实际负荷需求是未知的,提出一种基于 TD3 的异步耦合优化策略。为解决由预测误差、用户行为等引起的多重不确定性,采用了课程学习来设计“由易到难”的有逻辑训练过程,即通过在复杂且充满多重不确定性的系统环境中逐步增加学习难度,收敛速度提升了 21.88%。

1 市场机制及研究框架

本文假设一天被分成 T 个时刻,每个时刻用 t 表示。在典型的偏差结算电力批发市场中,售电商在日前电力批发市场投标购买日内每个时刻的电量,并在日内按照分时电价为多个用户提供电力供应和激励需求响应计划,日前市场购电与实际电力需求间的偏差在日内实时批发市场中结算。本文构建了售电商与多个用户的双层问题结构,如图 1 所示。售电商在批发市场出清电价高于售电价格的时段,制定需求响应策略以减少售电损失,而用户根据售电商制定的补贴决定相应时段响应量来获取收益。售电商与用户主从博弈架构属于双层问题结构。上层问题是售电商以自身全天累积利润最大为目标,优化电力批发市场的投标策略和发布给用户激励需求响应策略。下层问题是以用户每个时刻效用最大为目标优化用能行为。

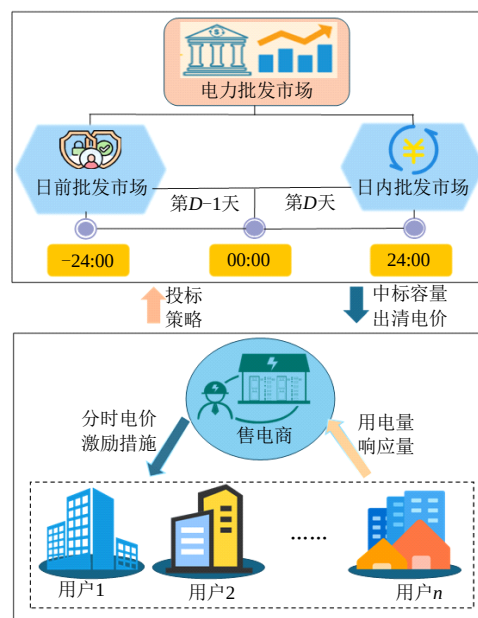


图 1 市场结构

Fig. 1 Market structure

在前一天,售电商作为独立个体向独立系统运营商(independent system operator, ISO)提交日内每个时间段的报价,ISO 根据批发市场的买方和卖方

的报价报量清算电力批发市场，以产生批发市场出清电价，以及每个参与者交易的电力。在日内时刻，售电商根据用户的负荷需求、投标所购电力、批发市场电价和分时电价向用户宣布激励需求响应策略。用户根据售电商宣布的激励策略来综合决策自己的用能行为，并向售电商反馈调整结果。当电力批发市场的出清价格高于分时电价时，向用户售卖电力会导致出现经济赤字，可以通过需求响应减少负荷需求来缓解。售电商通过从批发市场购买电力和向用户售电来获得收益。售电商承担价格风险，因为如果在前一天的市场上购买不足，则会出现短缺必须在日内批发市场购买，反之亦然。因此，售电商必须优化其在日前市场上的投标决策。

2 主体参与电力市场模型构建

2.1 售电商市场决策模型

售电商需要优化电力批发市场中的投标策略和电力零售市场中的需求响应激励策略。

根据微观经济学中的弹性理论，激励型需求响应的响应价格弹性 ε 可以定义为响应量对激励价格变化敏感度^[27]。由于用户舒适度损失直接影响到用户对激励价格变化的敏感度及响应程度， ε 随用户响应量 R 的增加而减小。由于用户的多样性， ε 在用户中的变化趋势是不同的。在现有的激励策略中，用户在激励型需求响应中的所有累积响应量都是以相同的激励价格结算的，这意味着 ε 的变化不能完全反映出来，也忽略了用户有限理性行为对激励效果的潜在干扰。为了充分利用用户的调节弹性差异与用户心理因素导致的非理性偏差，本文提出了一种考虑用户响应弹性与禀赋效应的增量激励需求响应策略，其中，售电商根据用户响应的增量给予相应的补偿，激励补偿成本为

$$C_{i,t}(\Delta Q_{i,t}) = \begin{cases} \int_0^{\Delta Q_{i,t}} u(x) dx, & \Delta Q_{i,t} > 0 \\ 0, & \text{其他} \end{cases} \quad (1)$$

式中： $\Delta Q_{i,t}$ 为时刻 t 中的第 i 个消费者的响应量； $u(x)$ 为增量激励函数，它与用户的响应水平呈正相关，单位激励成本随着用户的响应深度的增加而增加，时变增量激励函数的具体形式可以根据用户的需求响应特征及心理因素灵活选择。为保证在电力零售市场中不存在负激励，增量激励函数应为正且单调递增。此外，售电商会根据每个时间段的用户情况更新具体的激励函数的参数。由于本策略是数

据驱动的，其适用于边际响应成本递增的用户，主要针对的是售电商的用户群体中的居民用户和商业用户。

售电商向日前市场上报的信息为每个时刻 t 的投标点 $(P_{t,k}^{\text{bid}}, Q_{t,k}^{\text{bid}})$ ，其中， k 表示每个时刻 t 的投标点索引值，有 $P_{t,k}^{\text{bid}} > P_{t,k+1}^{\text{bid}}$ 及 $Q_{t,k}^{\text{bid}} < Q_{t,k+1}^{\text{bid}}$ 。时刻 t 的竞价策略为连接这些点的非递增分段曲线，如图2所示^[28]。售电商在时刻 t 内通过日前电力批发市场交易的购电量 Q_t^{p} 为

$$Q_t^{\text{p}} = Q_{t,k}^{\text{bid}} + (P_t^{\text{RT}} - P_{t,k}^{\text{bid}}) \frac{Q_{t,k}^{\text{bid}} - Q_{t,k+1}^{\text{bid}}}{P_{t,k}^{\text{bid}} - P_{t,k+1}^{\text{bid}}} \quad (2)$$

式中 P_t^{RT} 为电力批发市场的出清电价。

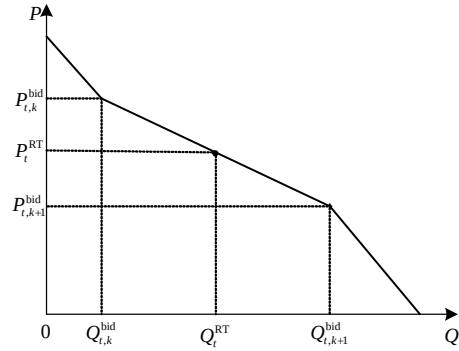


图2 售电商投标策略及日前批发市场购电数量

Fig. 2 Bidding strategy of electricity retailer and quantity purchased in the day-ahead wholesale market

售电商在时刻 t 在电力批发市场的购电成本 $C_{b,t}$ 为

$$C_{b,t} = P_t^{\text{RT}} Q_t^{\text{p}} \quad (3)$$

售电商需确保其提供的电力资源能够满足每个用户的电力负荷需求，应满足电力平衡的约束为

$$Q_t^{\text{p}} + Q_t^{\text{s}} = \sum_{i=1}^N Q_{i,t}^{\text{a}} \quad (4)$$

式中： $Q_{i,t}^{\text{a}}$ 为第 i 个消费者在时刻 t 中的实际负荷； Q_t^{s} 为时刻 t 中售电商日前购电导致日内为用户提供电力供应的电力偏差；假设有 N 个用户参与。

此外，任何时刻 t 中用户的总负荷不应超过输电线的最大容量 Q_{max} ，即 $0 \leq \sum_{i=1}^N Q_{i,t}^{\text{a}} \leq Q_{\text{max}}$ 。

售电商的优化目标是最大化全天多个时刻的累积利润：

$$\max_{u(x), (P_{t,k}^{\text{bid}}, Q_{t,k}^{\text{bid}})} \sum_{t=0}^{T-1} U_t = \sum_{t=0}^{T-1} \left(\sum_{i=1}^N C_{i,t}^{\text{p}} - \sum_{i=1}^N C_{i,t} - C_{b,t} \right) \quad (5)$$

式中： U_t 为售电商在时刻 t 中获得的利润； $C_{i,t}$ 为售

电商提供给 i 用户在时刻 t 中参与激励型需求响应的补偿成本； $C_{i,t}^p$ 为 i 用户在时刻 t 中在电力零售市场中向售电商的购电成本。

2.2 基于禀赋效应的有限理性用户模型

禀赋效应是行为经济学中的重要概念，指个体对已拥有的物品赋予高于其客观市场价值的主观价值评估，导致其对放弃该物品要求的补偿显著高于获取同类物品的支付意愿。这一现象主要由2种心理机制驱动：1) 损失厌恶，即用户对损失的敏感度显著高于同等收益；2) 心理账户，即用户倾向于将资源划分为不同的心理账户，并对不同账户赋予异质性价值^[29]。

传统经济学假设用户为完全理性决策者，能够基于成本-收益分析做出最优响应决策。然而，现实中的用户行为往往受到心理因素的影响，表现出有限理性。禀赋效应为解释用户参与需求响应中的非理性行为提供了重要视角。具体而言，用户对现有用电模式如固定用电时段、电器使用习惯赋予较高的心理价值，即使调整用电模式可带来经济收益，仍可能因禀赋效应选择维持现状。用户对需求响应中潜在的用电不便或不确定性表现出过度敏感，进一步凸显心理因素对决策的干扰^[30]。此外，用户的禀赋效应和损失厌恶系数受多种个体特征的影响，包括年龄、财富水平、文化背景和风险偏好等。这些特征会显著影响用户对资源的主观价值评估以及对损失的敏感度。在实践中，需要结合用户的具体特征和行为数据量化其禀赋效应特性，并充分考虑不同用户的有限理性行为。基于本文场景需要可简化表征，采用禀赋效应因子 $\mu_{i,k} > 1$ 反映用户放弃用电权的主观心理成本放大系数^[13]，将禀赋效应纳入到用户的行为模型中。

用户以参与需求响应的效用函数最大为目标：

$$\max U_{i,t}^{\text{user}} = C_{i,t}(\Delta Q_{i,t}) - \mu_{i,k} C_{i,t}^c(\Delta Q_{i,t}) - P_t^{\text{TOU}} Q_{i,t}^a \quad (6)$$

式中：第1项为用户 i 在时刻 t 的激励收益；第2项为用户 i 考虑禀赋效应的响应成本； $C_{i,t}^c(\Delta Q_{i,t})$ 为客观响应成本， $\mu_{i,k}$ 将其放大为考虑心理因素后的成本；第3项为用户 i 在时刻 t 中在电力零售市场中购电成本； P_t^{TOU} 为售电商在时刻 t 中事先定好的分时电价； $Q_{i,t}^a$ 为用户实际负荷需求。

用户在每个时刻都会有最舒适的用电量，即用户的基线负荷 $Q_{i,t}^b$ 。当用户的用电量偏离 $Q_{i,t}^b$ 时，会给用户造成舒适度损失。用户参与激励型需求响应

造成的舒适度损失与用户的响应和灵活性有关^[2]。

$$C_{i,t}^c(\Delta Q_{i,t}) = \frac{1}{2} \Delta Q_{i,t}^2 \theta_i + \lambda_i \Delta Q_{i,t} \quad (7)$$

式中： θ_i 为用户 i 的响应弹性倒数； λ_i 为用户 i 可以接受的最低激励价格。

传统用户行为模型仅考虑各个用户在当前时段内的最优响应策略，未能考虑当前时段的响应策略对未来时段响应策略的影响，即忽略了用户的动态响应特性。本文通过用户参与需求响应的负荷反弹效应表征解决用户的动态响应特性：

$$Q_{i,t}^a = Q_{i,t}^b - \Delta Q_{i,t} + \phi_{i,t,j}(\Delta Q_{i,j}) \quad (8)$$

式中 $\phi_{i,t,j}(\cdot)$ 为用户 i 在时刻 t 受时刻 j 负荷削减影响后的负荷反弹量，是历史时刻 j 的负荷削减量 $\Delta Q_{i,j}$ 的加权和。负荷反弹是指响应后一段时间内负荷需求高于正常负荷的现象，主要是由温控负载、电动汽车等可转移负荷引起的。负荷反弹量与响应的深度有关：

$$\phi_{i,t,j}(\Delta Q_{i,j}) = \sum_{j=0}^{t-1} \eta_{i,j,t} \Delta Q_{i,j} \quad (9)$$

式中 $\eta_{i,j,t}$ 为时刻 j 中负荷削减对时刻 t 负荷的影响系数，与历史时间段内负荷削减的累积效应密切相关。考虑到用户行为的不确定性会影响用户的动态响应特性，本文引入贝叶斯转移概率来描述用户行为的随机性，并反映用户群体行为对用户响应的动态累积效应，从而量化历史时刻负荷削减对当前时刻负荷反弹的动态影响。负荷反弹与需求响应之间的关系^[31]可以表示为

$$\eta_{i,j,t} = \frac{\sum_{j=0}^{t-1} \sum_{m=0}^M P_t\{Q_{i,t} = Q_{i,t-1} + \Delta Q_{m,j} \mid \Delta Q_{m,j}\} \Delta Q_{m,j}}{\sum_{j=0}^{t-1} \sum_{m=0}^M \Delta Q_{m,j}} \quad (10)$$

其中，假设在时刻 j 中有 M 个用户参与需求响应，其第 m 个用户的响应量为 $R_{m,j}$ ，第 m 个用户负荷削减量转移到时刻 t 的概率可以表示为：

$P_t\{Q_{i,t} = Q_{i,t-1} + \Delta Q_{m,j} \mid \Delta Q_{m,j}\}$ 。分子表示所有参与响应的用户在历史时刻 j 的负荷削减 $\Delta Q_{m,j}$ 通过贝叶斯转移概率加权求和，表示这些负荷削减对当前时刻 t 负荷状态的预期影响。分母表示所有响应的用户在历史时刻的总负荷削减量，用于归一化影响系数 $\eta_{i,j,t}$ ，其中，响应量 $\Delta Q_{i,t}$ 满足：

$$0 \leq \Delta Q_{i,t} \leq Q_{i,t}^b + \phi_{i,t,j}(\Delta Q_{i,j}) \quad (11)$$

3 基于禀赋效应因子的增量激励策略优越性分析

3.1 考虑禀赋效应的激励策略动态匹配

考虑禀赋效应后用户的边际成本满足：

$$z(\Delta Q_{i,t}) = \mu_{i,k} \frac{\partial C_{i,t}^C}{\partial R} = \mu_{i,k} (\theta_i \Delta Q_{i,t} + \lambda_i) \quad (12)$$

增量激励策略中，激励函数 $u(\Delta Q)$ 需动态跟踪用户边际成本的变化，用户的边际激励补偿理论上需满足：

$$u(\Delta Q_{i,t}) = \mu_{i,k} (\theta_i \Delta Q_{i,t} + \lambda_i) \quad (13)$$

表明禀赋效应因子 $\mu_{i,k}$ 直接影响单位激励价格的设计。若忽略则激励补偿将低估用户心理成本，导致用户响应量不足。由于禀赋效应因子 $\mu_{i,k}$ ，用户的边际成本曲线上移，导致等价激励策略在相同激励水平下响应量减少，而增量激励策略可以通过动态调整 $u(\Delta Q)$ ，补偿禀赋效应带来的心理损失，维持用户响应深度。由于禀赋效应降低了用户对激励价格的敏感度，需通过动态调整激励函数来补偿心理成本。

3.2 考虑禀赋效应后的消费者剩余再分配

当用户通过激励效益与负荷削减成本的净收益即消费者剩余最大化时，用户会选择最优响应量进行决策：

$$\Delta Q^* = \operatorname{argmax} \int_0^{\Delta Q} [u(x) - z(x)] dx \quad (14)$$

式中： $z(x)$ 为用户边际成本函数(含实际成本与禀赋效应心理成本)； $\int_0^{\Delta Q} [u(x) - z(x)] dx$ 为用户的消费者剩余。

现有激励策略中等价激励函数为 $g(\Delta Q)$ ，其根据用户在特定时刻的响应总量确定。假设在特定的激励型需求响应事件中，2 种策略下激励函数和低灵活性响应函数相交于同一点，即 $\Delta Q_{1,u} = \Delta Q_{1,g}$ ， $u(\Delta Q_{1,u}) = g(\Delta Q_{1,g})$ 。由于 $u(\Delta Q)$ 为单调递增函数，当 $\forall \Delta Q_{1,u} = \Delta Q_{1,g}$ 时，增量激励成本始终低于固定价格的等价激励成本，即有：

$$C_u(\Delta Q_{1,u}) = \int_0^{\Delta Q_{1,u}} u(x) dx < C_g(\Delta Q_{1,g}) \quad (15)$$

由此表明：当消费者的响应相同时，增量激励策略的成本总是低于现有的等价激励策略。增量激励策略能够更好地跟踪用户边际成本变化，使用户每单位响应所获得的消费者剩余在响应累积过程中保持在合理的范围内，从而降低售电商的激励

成本。

接着，分析 2 种不同灵活性用户在 2 种激励策略下的响应情况。考虑禀赋效应后用户在决策时会考虑放弃现有用电模式的心理成本。低灵活性用户由于自身响应程度有限，在 2 种激励策略下所能决策的最大响应量相同。高响应弹性用户在决策时，会综合权衡激励收益与包括心理成本在内的总成本。假设低响应弹性用户和高响应弹性用户的响应函数分别表示为 $I_1 = z_1(\Delta Q)$ 和 $I_2 = z_2(\Delta Q)$ 。

由 $g(\Delta Q) = z_2(\Delta Q)$ 可获得高灵活用户在激励策略的响应量 $\Delta Q_{2,g}$ 。由于增量激励函数是单调递增的，并且当 $\Delta Q_{1,u} = \Delta Q_{1,g}$ 时，考虑禀赋效应使得用户心理成本增加，在相同激励下，用户更倾向于维持现有用电模式，所以，对于高响应弹性用户而言可得：

$$u(\Delta Q) > g(\Delta Q), \forall \Delta Q > \Delta Q_{1,g} \quad (16)$$

由增量响应所产生的消费者盈余的变化为

$$\frac{d \int_0^{\Delta Q} [u(x) - z_2(x)] dx}{d \Delta Q} \Big|_{\Delta Q = \Delta Q_{2,g}} = u(\Delta Q_{2,g}) - z_2(\Delta Q_{2,g}) > 0 \quad (17)$$

可知，高灵活用户在增量激励策略中在 $\Delta Q_{2,g}$ 处增加响应深度可增加用户收益，高响应弹性消费者将增加响应，直到消费者剩余的增量为 0。

由此求解 $u(\Delta Q) = z(\Delta Q)$ 获得高响应弹性用户的响应量 $\Delta Q_{2,u}$ ，高灵活用户在增量激励策略下响应量大于等价激励策略下的响应量，可得：

$$\frac{\Delta Q_{2,u}}{\Delta Q_{1,u}} > \frac{\Delta Q_{2,g}}{\Delta Q_{1,g}}, \forall \Delta Q_{1,u} = \Delta Q_{1,g} \quad (18)$$

结果表明，考虑禀赋效应后增量激励策略相比于等价激励策略能增加高灵活用户的响应深度并将更多的消费者剩余转移给高响应弹性用户。受禀赋效应影响，高响应弹性用户对激励变化更敏感，增量激励策略能更好契合其心理预期，增强其参与需求响应的积极性。通过消费者剩余的再分配，增量激励策略在用户中实现了激励补偿与心理损失的动态平衡，避免了等价策略下的分配不公。

4 优化问题架构及求解

4.1 基于马尔可夫决策过程的优化决策问题构建

一个标准的马尔可夫决策过程 (Markov decision process, MDP)^[32] 可以用来描述异步耦合优化问题，被定义为一个五元组 (S, a, P, R, γ) ，其中，

$\mathcal{S}$ 表示状态空间(state space); $\mathcal{a}$ 表示动作空间(action space); P 表示状态转移概率(state transition probability); R 表示奖励函数(reward function); γ 表示折扣因子。

4.1.1 状态空间

为优化售电商在日前电力批发市场的投标策略, 需综合考虑日前电力批发市场预测出清电价 $\tilde{P}_t^{\text{RT}}$ 、分时电价 P_t^{TOU} 、上一时刻批发市场出清电价 P_{t-1}^{RT} 、日内未实施激励需求响应的负荷需求量 $Q_{i,t}^{\text{b}}$ 及用户日内实际负荷需求 $\tilde{Q}_{i,t}^{\text{a}}$ 。因此, 售电商日前优化投标策略时的状态空间 $\mathcal{S}_t^{\text{DA}}$ 表示为

$$\mathcal{S}_t^{\text{DA}} = (\tilde{P}_t^{\text{RT}}, P_{t-1}^{\text{RT}}, P_t^{\text{TOU}}, Q_{i,t}^{\text{b}}, \tilde{Q}_{i,t}^{\text{a}}) \quad (19)$$

为优化售电商在日内实施的激励需求响应策略, 需综合考虑电力批发市场中的出清电价 P_t^{RT} 、分时电价 P_t^{TOU} , 预测日内未进行激励需求响应的负荷需求 $Q_{i,t}^{\text{b}}$ 、禀赋效应因子 $\mu_{i,k}$ 及负荷削减影响系数 $\eta_{i,j,t}$ 。因此, 售电商日内优化增量激励需求响应策略时的状态空间 $\mathcal{S}_t^{\text{RT}}$ 表示为

$$\mathcal{S}_t^{\text{RT}} = (P_t^{\text{RT}}, P_t^{\text{TOU}}, Q_{i,t}^{\text{b}}, \mu_{i,k}, \eta_{i,j,t}) \quad (20)$$

4.1.2 动作空间

由售电商执行的动作包含日前市场投标策略相关的每个时刻 t 的竞价点 $(P_{t,k}^{\text{bid}}, Q_{t,k}^{\text{bid}})$ 以及与增量激励函数 $u(x)$ 相关的所有参数 π_t 。售电商的投标策略在日前完成, 增量激励策略决策是在日内完成。动作空间分为日前动作空间和日内动作空间, 分别表示为:

$$\mathcal{a}_t^{\text{DA}} = [(P_{t,k}^{\text{bid}}, Q_{t,k}^{\text{bid}}), k=0,1,2,3] \quad (21)$$

$$\mathcal{a}_t^{\text{RT}} = (\pi_t) \quad (22)$$

值得注意的是, 需要优化决策的参数 π_t 的个数由满足要求的增量激励函数 $u(x)$ 的形式决定。

4.1.3 奖励函数

为了优化售电商全天售电累积利润, 由于日前投标智能体与激励成本无直接关联, 其优化投标策略的目标是降低因购电偏差导致的利润损失及提高购售电收益, 其奖励函数 $r_{\text{DA},t}$ 设置为

$$r_{\text{DA},t} = \sum_{i=1}^m C_{i,t}^{\text{p}} - C_{\text{b},t} - D_{\text{s},t} \quad (23)$$

式中 $D_{\text{s},t}$ 为售电商因日前投标购电偏差产生的惩罚项。考虑到本文采用数据驱动算法, 其具体形式不影响求解, 本文将其表示为

$$D_{\text{s},t} = \begin{cases} 2P_t^{\text{RT}}Q_t^{\text{s}}, & Q_t^{\text{s}} > 0 \\ 0.5P_t^{\text{RT}}Q_t^{\text{s}}, & Q_t^{\text{s}} < 0 \end{cases} \quad (24)$$

本文出于优化售电商的全天售电累积利润, 由于日内激励策略智能体的目标是优化决策激励策略以降低激励成本及提高购售电收益, 其奖励函数设置为

$$r_{\text{RT},t} = \sum_{i=1}^m (C_{i,t}^{\text{p}} - C_{i,t}) - C_{\text{b},t} \quad (25)$$

累积奖励 R_t 表示为

$$R_t = \sum_{k=t}^T \gamma^{k-t} r_k \quad (26)$$

式中 γ 为折扣因子。

4.2 基于课程学习的 TD3 异步耦合优化策略

售电商需依次在日前电力批发市场中观测 $\mathcal{S}_t^{\text{DA}}$ 做出投标策略, 在日内电力零售市场中观测 $\mathcal{S}_t^{\text{RT}}$ 实施激励策略。由于售电商在制定投标策略时需知晓用户因进行激励需求响应后的实际负荷需求, 2 个策略是耦合的。该异步耦合决策问题可以转化为 2 层优化问题, 上层是投标问题, 下层是激励需求响应问题。优化下层问题时需要上层的电力批发市场出清电价和投标购得电量作为输入。下层的用户响应量也会影响上层的投标决策。然而, 投标策略和激励需求响应策略是按顺序依次执行的, 这将导致售电商不能提前知晓用户进行激励需求响应后的实际负荷需求。在训练智能体时, 奖励函数是根据投标策略与激励需求响应策略两者优化的共同结果计算得出的, 因此, 智能体执行投标动作后无法立即获得奖励。

由于多样化用户模型参数在实践中难以准确识别, 当前研究在面对无模型环境时, 通过 DRL 和启发式算法来寻找最优解。然而, 售电商在进行投标决策时, 需求响应的行为和优化结果往往是未知的。由于在线启发式算法依赖于完整的信息环境, 其无法有效求解该问题。鉴于神经网络学习和记忆能力, 深度强化学习能够在信息不完全的情况下推断激励需求响应的潜在优化结果, 从而制定最优的投标策略。此外, 相比于传统的离线深度强化学习算法(例如深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法), TD3 通过引入双重估计、策略延迟更新以及目标策略平滑等技术, 可以提升算法的性能和稳定性^[33]。然而, TD3 在训练过程中依然面临样本效率和收敛速度

等挑战。

需求侧用电量的随机性、现货批发市场出清电价的波动性及受负荷反弹影响的动态响应不确定性显著增加了售电商优化策略训练的复杂性。为了提高模型的训练效率，本文将课程学习嵌入到售电商训练模型中。课程学习是一种机器学习策略，通过分阶段、由易到难的策略，提高求解效率和最终结果的精度。本文设计了一种基于训练次数的 3 阶段课程策略，以加快智能体的学习速度^[34]。具体来说，在第 1 阶段，构建一个理想化的学习环境，假设用户的用电行为、现货市场出清电价及负荷反弹系数均为确定性条件；在第 2 阶段，引入用户用电行为和现货市场出清电价的的部分不确定性，增加学习任务的复杂性，使智能体能够适应一定程度的不

确定性；在第 3 阶段，在高度不确定性的复杂环境下训练智能体，使其能够处理异步耦合优化策略。

设 G_{DA} 和 G_{RT} 分别代表日前投标策略和日内激励策略的智能体。TD3 是一种离线学习算法，虽然动作是按时间序列执行的，但在网络训练是独立的。由于日前投标时用户的实际负荷需求会受到激励需求响应的影响，所需购电量是未知的，因此先对日内激励策略智能体进行训练，然后根据日内智能体激励策略 π_{RT} 推测出用户因参与需求响应后实际用电需求 $\tilde{Q}_{i,t}^a$ ，即 $Q_{i,t}^b \xrightarrow{\pi_{RT}} \tilde{Q}_{i,t}^a$ 。故将日内激励策略智能体作为环境参与日前投标策略智能体的训练。值得注意的是，由于首先训练日内激励策略智能体，因此，训练时假设日前购买电量都能满足日内的需求。所提出的算法的框架如图 3 所示。

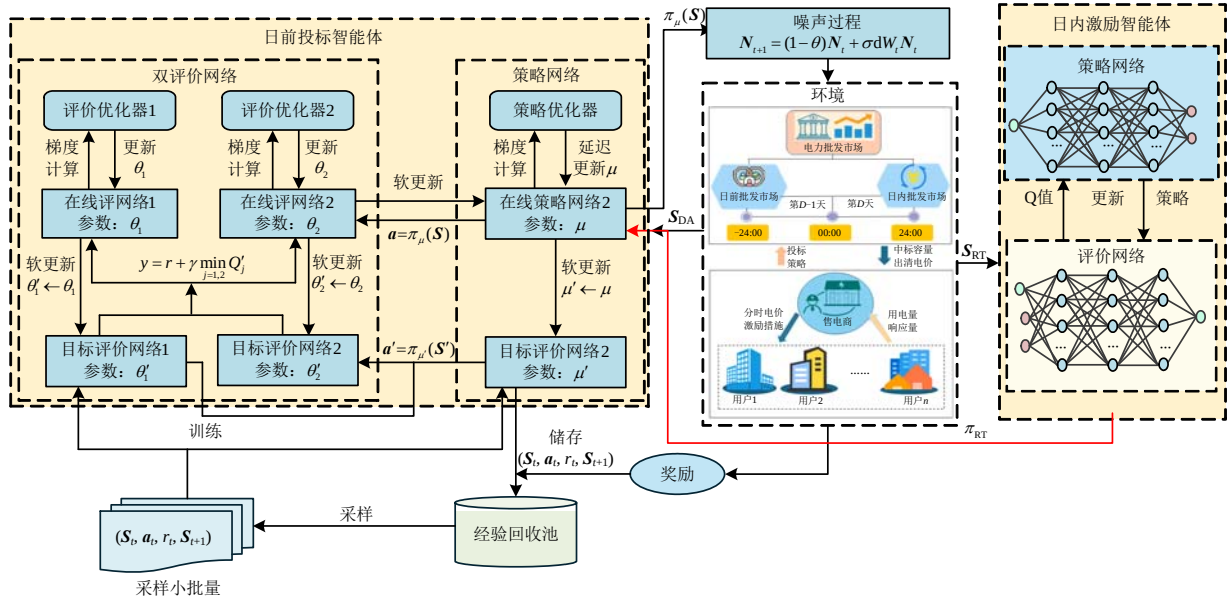


图 3 基于 TD3 的异步耦合优化方法框架

Fig. 3 Framework of asynchronous coupling optimization method based on TD3

TD3 算法的目标是学习好最佳策略 π ，即有：

$$\max_{\pi} J[\pi] = \mathbb{E}_{a_i \pi} [R'_i] \quad (27)$$

式中： $J[\pi]$ 为目标函数； $\mathbb{E}$ 为数学期望。

为了衡量策略 π 的性能，本文采用基于贝尔曼方程的价值函数，即有：

$$Q^{\pi}(S, a) = \mathbb{E}\{r + \gamma Q^{\pi}(S', \pi(S')) | S, a\} \quad (28)$$

式中： $Q^{\pi}(S, a)$ 为当前时段的每个动作的总期望返回量； $\pi(S')$ 为动作函数，可将状态映射到特定动作来指定策略。

设 $\theta^{Q_{1,2}}$ 和 θ^{μ} 分别表示评价网络和策略网络的参数向量， $\theta^{Q_{1,2}'}$ 和 $\theta^{\mu'}$ 分别表示目标评价网络和目标策略网络的参数向量。最小化损失函数来更新评

价网络的参数向量 θ^{Q_j} ：

$$L(\theta^{Q_j}) = \sum_i [y_i - Q(S_i, a_i | \theta^{Q_j})]^2 / H \quad (29)$$

式中 H 为从经验回放池中提取的样本数量。为了缓解价值函数高估偏差的问题，并在训练过程中来提高策略的稳定选性，TD3 采用 2 个评价网络，并择最小值作为时序差分目标，采用裁剪双 Q 学习来更新 Q 值。 y_i 使用目标网络计算， $y_i = r_i + \gamma \min_{j=1,2} Q'_j[S'_{i+1}, \mu'(S'_{i+1} | \theta^{\mu'}) | \theta^{Q'_j}]$ 。

通过策略梯度对策略网络的参数向量 θ^{μ} 进行更新，即有：

$$\nabla_{\theta^{\mu}} J \approx \frac{1}{H} \sum_i \nabla_a Q(S, a | \theta^Q) |_{S=S_i, a=\mu(S_i)} \nabla_{\theta^{\mu}} \mu(S | \theta^{\mu}) |_{S_i} \quad (30)$$

式中: ∇_{θ^μ} 为 θ^μ 的梯度; ∇_a 为 a 的梯度。

评价网络参数向量更新方式为

$$\theta^{Q'} = \theta^Q - \nabla_{\theta^Q} L(\theta^Q) \quad (31)$$

策略网络参数 $\theta^{\mu'}$ 更新方式为

$$\theta^{\mu'} = \theta^\mu + \nabla_{\theta^\mu} J(\theta^\mu) \quad (32)$$

由于 TD3 是确定性策略,为提高动作空间的探索效率,通过将噪声过程采样的噪声添加到动作策略中,构造了探索策略 μ' , 即有:

$$\mu'(S_t) = \mu(S_t | \theta^\mu) + N_t \quad (33)$$

式中噪声过程 N_t 采用奥恩斯坦-乌伦贝格过程^[35]。

基于 TD3 的异步耦合优化策略训练过程如下:

1) 随机初始化智能体 G_{DA} 的评价网络参数 θ_{DA}^Q 、 θ_{DA}^Q 和策略网络参数 θ_{DA}^μ 及智能体 G_{RT} 的评价网络参数 θ_{RT}^Q 、 θ_{RT}^Q 和策略网络的参数 θ_{RT}^μ ;

2) 初始化经验回放池 R_{DA} 、 R_{RT} ;

3) 回合数=1, ..., M;

4) 初始化随机过程, N_{DA} 选择 a^{DA} , N_{RT} 选择 a^{RT} ;

5) 接收初始状态空间 S_t^{DA} 和 S_t^{RT} 时刻 $t=0, \dots, T-1$;

6) 选择 $a_t^{DA} = \mu(S_t^{DA} | \theta_{DA}^\mu) + N_{DA,t}$;

7) 日前电力批发市场中执行 a_t^{DA} ;

8) 日内电力零售市场中执行 a_t^{RT} 并接收 r_t 和 S_{t+1} , 经验回收池 r_{DA} 、 r_{RT} 中存储 $(S_t^{DA}, a_t^{DA}, r_t, S_{t+1}^{DA})$ 、 $(S_t^{RT}, a_t^{RT}, r_t, S_{t+1}^{RT})$;

9) 如果激励策略训练尚未完成, 从经验回收池 R_{RT} 中随机抽样 H_{RT} 个样本, 通过式(31)更新 G_{RT} 评价网络, 通过式(32)更新 G_{RT} 策略网络, 更新 G_{RT} 的目标策略网络和评价网络为 $\theta_{RT}^{Q_{1,2}} \leftarrow \tau_{RT} \theta_{RT}^{Q_{1,2}} + (1 - \tau_{RT}) \theta_{RT}^{Q_{1,2}}$ 、 $\theta_{RT}^{\mu'} \leftarrow \tau_{RT} \theta_{RT}^{\mu'} + (1 - \tau_{RT}) \theta_{RT}^{\mu'}$;

10) 否则, 从经验回放池中随机抽取少量样本的过渡;

11) 根据得到激励需求响应结果获得状态空间 S_t^{DA} ;

12) 通过式(31)更新 G_{DA} 评价网络, 通过式(32)更新 G_{DA} 策略网络. 更新 G_{DA} 的目标策略网络和评价网络为 $\theta_{DA}^{Q_{1,2}} \leftarrow \tau_{DA} \theta_{DA}^{Q_{1,2}} + (1 - \tau_{DA}) \theta_{DA}^{Q_{1,2}}$ 、 $\theta_{DA}^{\mu'} \leftarrow \tau_{DA} \theta_{DA}^{\mu'} + (1 - \tau_{DA}) \theta_{DA}^{\mu'}$ 。

5 算例分析

5.1 算例设置

本文所采用的电力批发市场日前出清电价取

自 PJM 电力市场。假设有 25 位电力用户, 其基线负荷数据来自 Pecan Street 数据库^[36]。分时电价是根据文献[37]中研究结果设定的。增量激励函数可定义为 $u(R) = \alpha_1 + \beta R$, 等价激励价格定义为 $g(R) = \alpha_2$ 。用户模型参数 θ_i 及 λ_i 分别在 $[0, 1.5]$ 和 $[0, 0.7]$ 内随机生成。禀赋损失厌恶系数取自参考文献[13]。假设负荷反弹受过去时间段的影响, 由于不确定性的影响, 负荷反弹系数服从正态分布, 标准差为 0.01。为了模拟投标策略优化时由预测误差引起的不确定性, 假设在实际值中添加一个干扰来模拟日前出清电价预测值和实际负荷需求预测值, 该干扰均服从正态分布, 均值为 0, 标准差为 0.05。批发市场出清电价、用户基线负荷需求以及用户响应量等不确定性均服从均值为 0、标准差为 0.03 的正态分布。所有不确定性可表示为 $X_u = X_0[1 + N(e, \sigma^2)]$, 其中, X_0 为无不确定性的原始值; X_u 为有不确定性的值; $N(e, \sigma^2)$ 表示均值为 e 、标准差为 σ 的正态分布。

本文使用 93 天的数据作为训练集用于训练模型; 使用 10 天的数据作为验证集; 使用 2 天的数据作为测试集, 用于测试模型训练的效果。数值验证模拟在 2 天内进行, 一天被分为 24 个时刻, 即时刻 $t=1h$ 。采用的强化学习算法的所有网络采用具有 3 层隐藏层的全连接网络, 其中, 前 2 层包含 256 个神经元, 第 3 层包含 128 个神经元。折扣因子设置为 0.99。2 个子代理(Actor 和 Critic 网络)的学习率分别设定为 0.000 001 和 0.000 01。

5.2 所提策略性能分析

图 4 显示由于现货市场价格波动, 可能出现日前出清电价高于分时电价的情况(即“价格倒挂”)。此时, 售电商通过增量激励策略动态匹配用户的边际成本与禀赋效应心理损失, 显著减少用户负荷需求。例如, 在第 17 时刻, 购电价格约 1.74 元/(kW·h),

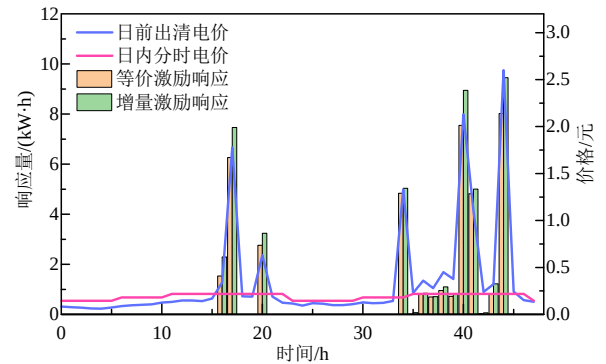


图4 在所选2天中2种激励策略下的响应及电价

Fig. 4 Response and electricity price under two incentive strategies in the selected two days

售电商每出售 1 度电将要亏损 1.52 元, 增量激励策略的响应量为 7.4 kW·h, 而等价激励策略为 6.3 kW·h, 总响应量较等价策略提升约 17.5%。第 40 时刻的购电价格为 2.10 元/(kW·h), 售电亏损达 1.89 元/(kW·h), 增量激励策略通过动态匹配用户边际成本与禀赋效应心理损失, 吸引高响应弹性用户深度挂程度增加, 增量策略通过动态调整激励函数匹配边际成本变化, 从而提升响应效率。随着电价倒挂程度增加, 售电商可适度提高激励成本以进一步激发用户响应潜力, 减少用电需求。相反, 在日前电价低于分时电价时, 售电商无需实施需求响应策略。

图 5 展示了出现价格倒挂的典型时刻的 2 种激励策略下需求响应的单位成本和响应情况。值得注意的是, 该图的轴线的间隔并不代表等量的时间间隔。柱状图表示售电商在 2 种激励策略中的平均单位激励成本, 点线图表示所有用户的总响应量。从第 17 时刻可知, 增量策略单位激励成本为 1.08 元/(kW·h), 等价策略为 1.31 元/(kW·h), 并且增量策略响应量更高。这表明增量策略通过跟踪用户响应弹性, 避免了响应初期的消费者剩余过度分配, 同时禀赋效应因子的引入确保了激励补偿与用户心理成本动态匹配, 进一步缓解响应初期过度的消费者剩余, 降低了单位激励成本。从图 5 其他时刻也同样可看出, 增量激励策略的响应量高于等价激励策略, 同时, 其平均单位激励成本低于等价激励策略的单位激励成本。这表明, 计及禀赋效应后的增量激励策略比等价激励策略具有优越性。这是因为等价激励策略总是与用户总响应量耦合, 采用的固定激励水平在响应初期消费者盈余较高, 导致单位激励成本较高。增量激励策略利用禀赋效应因子与用户的响应弹性通过调整激励水平来动态匹配用户

边际成本与禀赋效应心理损失, 优化激励资源分配, 降低了单位激励成本。在第 16、35—39 及 42—43 时刻出清电价略低于分时电价, 此时造成的经济赤字较小。售电商为降低售电损失只能提供较低的激励补贴, 用户响应量较小, 2 种策略的激励水平差异尚未明显显现。

在第 44 时刻中, 增量策略单位成本为 1.60 元/(kW·h)略高于等价策略, 但增量策略响应量为 8.1 kW·h, 等价策略响应量为 6.8 kW·h, 响应量的提升显著缓解了售电商的经济赤字。售电商在增量激励策略的表现依旧优于等价激励策略。这是因为高弹性用户在增量策略下因边际激励递增而更愿意深度响应。这是由于增量激励策略给予用户的每单位激励补偿是基于响应增量而增加, 其获得的单位响应激励补贴价格也在增加。增量激励策略吸引更多高响应弹性用户参与响应, 虽然激励支出增加, 但负荷响应效果显著优于等价激励, 其可更多地减少因价格倒挂导致的经济赤字。在所提策略中售电商因用户响应深度增加需更多激励成本, 与此同时, 却能更多地缓解因售电造成的经济赤字问题。虽然增量激励在某些时段的单位激励成本略高, 但其更高的负荷削减效果和整体效率使其成为更适合需求响应的激励策略。

如图 6 所示, 每条虚线表示每个用户考虑禀赋效应后的负荷响应函数, 指的是有限理性用户单位激励与负荷响应量间的关系。图 6 中的柱状图显示售电商制定的当前时刻激励策略。图 6 通过用户负荷响应函数(虚线)与激励价格(柱状图)的差值即消费者剩余来分析策略差异。

由图 6 可知, 若售电商为追求更多负荷削减而提高单位激励补贴, 则等价激励策略会因用户禀赋效应加剧响应初期消费者剩余的分布不均衡。在第 20 时刻, 等价策略下消费者剩余高达 0.66 元/(kW·h), 主要集中在响应初期; 而增量激励策略下消费者剩余为 0.10 元/(kW·h), 显著低于等价激励策略。这是因为增量激励通过动态调整单位激励价格, 逐步补偿用户边际成本与禀赋效应心理损失, 避免初期固定激励价格与用户心理成本脱节, 从而减少低弹性用户的超额剩余。在第 44 时刻, 增量激励策略的消费者盈余为 0.12 元/(kW·h), 而相应的等价策略的消费者盈余为 0.96 元/(kW·h)。通过时变激励函数, 随响应量递增动态匹配用户边际成本与禀赋效应损失, 避免低弹性用户初期盈

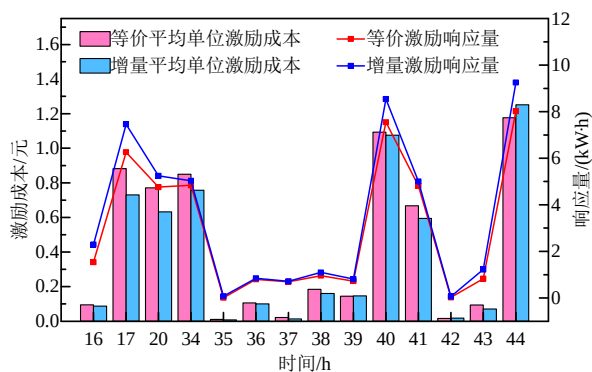


图 5 2 种策略下价格倒挂时单位平均激励成本及响应
Fig. 5 Unit average incentive cost and response at the time of price inversion under two strategies

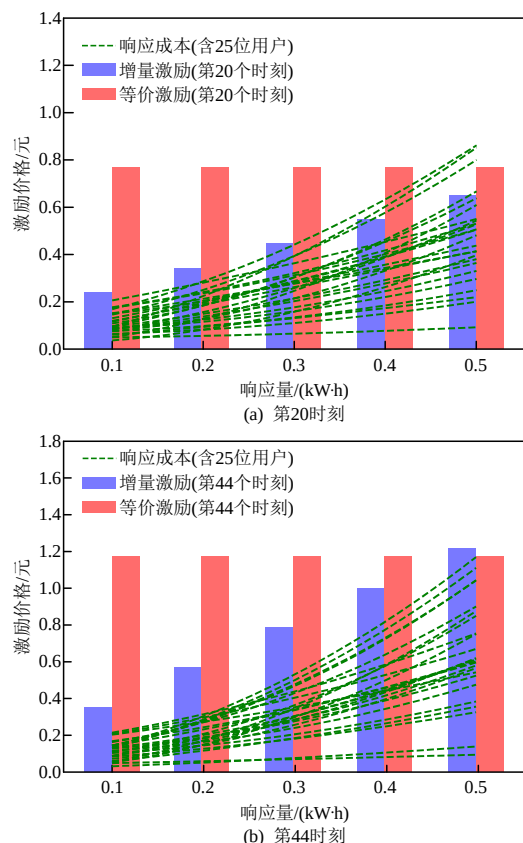


图6 典型时刻的2种激励策略及用户响应曲线

Fig. 6 Two incentive strategies and user response curves at typical moments

余过高。在价格倒挂加剧时,高弹性用户因边际激励递增更愿意深度响应,而禀赋效应因子放大了用户对用电权的主观估值,增量策略通过跟随用户边际成本的变化动态匹配心理成本,减少剩余积累,实现剩余从低弹性向高弹性用户转移,提升整体效率。消费者剩余的重新分配提高了售电商的激励效率,使其能够以更低的单位激励成本获得更多的负荷削减。

5.3 所提算法性能分析

为了验证所提出的基于课程学习的改进 TD3 算法的有效性,选择与传统的 TD3 及 DDPG 算法进行比较。

图7对比了3种算法在75000次训练回合的累积奖励曲线。值得注意的是,图7展示的曲线进行了平滑处理,1个回合设置为24个时间步,1个时间步为1h。从图7可以看出,由于智能体与环境的交互,累积奖励值随着训练回合数的增加而逐渐上升,表明算法在训练过程中逐渐学习到更优的策略。随着时间的推移,累积奖励值会逐渐稳定下来,表明智能体已能够在不确定性环境状态下做出每个时间步的最优激励策略与投标决策,以获得较高

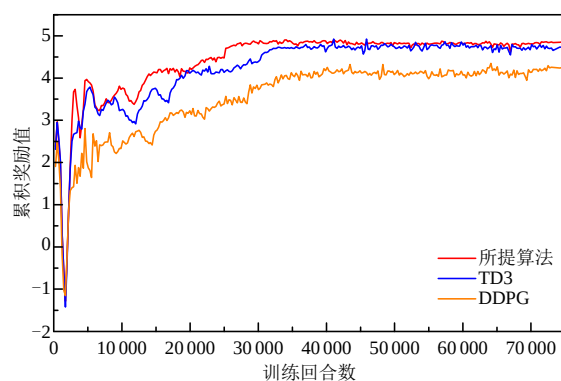


图7 对比算法及所提算法的训练结果

Fig. 7 Training results of compare algorithms and the proposed algorithm

的累积奖励。

从图7可知,所提算法通过课程学习逐步适应多重不确定性,确保策略在动态环境下的鲁棒性。在高不确定性下的第3阶段训练时算法仍能稳定收敛,验证了策略对禀赋效应与动态响应的双重适应性。具体来说,所提算法大约在25000回合奖励值收敛到4.8左右,传统的TD3在32000回合收敛到4.7左右,所提算法的收敛速度比传统TD3提升了约21.88%。结果表明,课程学习可以提高优化算法的训练效率,解决了优化方法在多方面不确定环境中收敛速度慢的问题。采用的DDPG算法大约在34000次回合训练后奖励值收敛到4.2左右。这是因为TD3相较于DDPG可以有效缓解Q值过估计及降低神经网络的更新频率。因此,本文所提的方法不仅在收敛速度方面具有优势,最终性能也明显优于其他2种算法。

6 结论

针对现货市场价格波动及用户有限理性行为对需求响应策略的挑战,从零售商的视角出发,本文提出一种考虑用户响应弹性差异与禀赋效应的增量激励策略。该策略根据响应增量分配激励补贴,可更合理地分配消费者盈余,有效解决了响应初期的消费者剩余过度问题。本文将用户禀赋效应及动态响应特性纳入用户响应模型中,考虑了其对增量激励策略的影响。通过理论分析验证了策略在激励资源再分配、降低售电商成本及提升高弹性用户响应深度方面的有效性。针对售电商决策的复杂性,构建了一个日前投标策略与日内激励策略的两阶段异步耦合优化问题,并提出基于课程学习的改进TD3异步耦合优化方法,高效地解决了该两阶段异步耦合决策问题中的信息未知问题。

本文解决了单一零售商激励多用户响应初期消费者剩余过度的问题。实际中, 单个用户可能同时参与多个售电商的需求响应项目, 进而引发售电商间的竞争或合作。未来研究可进一步考虑用户跨多售电商参与的复合行为特征, 深入探索售电商间的竞争或合作关系, 并刻画这种多主体动态博弈对激励策略设计、用户响应行为及整体市场效率的动态影响机制。

参考文献

- [1] 成润坤, 王辉, 刘达, 等. 融合 RSDE 框架与 KAN 算法的现货电价异质模型集成预测[J]. 中国电机工程学报, 2024, 44(24): 9645-9657.
CHENG Runkun, WANG Hui, LIU Da, et al. Ensemble prediction of spot electricity prices using heterogeneous models by integrating the RSDE framework and KAN algorithm[J]. Proceedings of the CSEE, 2024, 44(24): 9645-9657(in Chinese).
- [2] 郭昆健, 高赐威, 林国营, 等. 现货市场环境下售电商激励型需求响应优化策略[J]. 电力系统自动化, 2020, 44(15): 28-35.
GUO Kunjian, GAO Ciwei, LIN Guofang, et al. Optimization strategy of incentive based demand response for electricity retailer in spot market environment [J]. Automation of Electric Power Systems, 2020, 44(15): 28-35(in Chinese).
- [3] GUO Zhilong, XU Wei, YAN Yue, et al. How to realize the power demand side actively matching the supply side? —A virtual real-time electricity prices optimization model based on credit mechanism[J]. Applied Energy, 2023, 343: 121223.
- [4] 任恒宇, 韩冬, 任曦骏, 等. 基于时变价格弹性矩阵的深谷电价多目标定价策略[J]. 电网技术, 2024, 48(3): 958-967.
REN Hengyu, HAN Dong, REN Xijun, et al. Multi-objective deep valley electricity pricing model based on time-varying price elasticity matrix[J]. Power System Technology, 2024, 48(3): 958-967(in Chinese).
- [5] MEI Shufan, TAN Qinliang, LIU Yuan, et al. Optimal bidding strategy for virtual power plant participating in combined electricity and ancillary services market considering dynamic demand response price and integrated consumption satisfaction[J]. Energy, 2023, 284: 128592.
- [6] ZHEN Cheng, NIU Jide, TIAN Zhe, et al. Risk-averse transactions optimization strategy for building users participating in incentive-based demand response programs[J]. Applied Energy, 2025, 380: 125009.
- [7] 郑颖颖, 缪新义, 王晖, 等. 支持时变需求响应激励价格的电动汽车优化充电策略[J]. 电力系统自动化, 2025, 49(9): 96-106.
ZHENG Yingying, MIAO Xinyi, WANG Hui, et al. Optimal charging strategies for electric vehicles supporting time-varying demand response incentive price [J]. Automation of Electric Power Systems, 2025, 49(9): 96-106(in Chinese).
- [8] 戴逢哲, 姜飞, 陈磊, 等. 价格-积分联合激励下考虑消费舒适度的居民需求响应优化策略[J]. 电网技术, 2024, 48(2): 819-829.
DAI Fengzhe, JIANG Fei, CHEN Lei, et al. Optimization strategy of residential demand response considering consumption comfort Under Price-score joint incentive [J]. Power System Technology, 2024, 48(2): 819-829(in Chinese).
- [9] WANG Yunchu, YAN Yong, LIN Zhenzhi, et al. Score-based incentive demand response for load aggregators considering power-score redeem behavior of residential consumers[J]. International Journal of Electrical Power & Energy Systems, 2024, 162: 110233.
- [10] 吕挺, 张智, 芦鹏飞, 等. 考虑用户信用的虚拟聚合商需求响应交易机制[J]. 电力系统自动化, 2023, 47(19): 44-52.
LV Ting, ZHANG Zhi, LU Pengfei, et al. Demand response transaction mechanism for virtual load aggregator considering user credit[J]. Automation of Electric Power Systems, 2023, 47(19): 44-52(in Chinese).
- [11] QIAN Liang, LIN Shunfu, LI Fangxing, et al. Low carbon optimization dispatching of energy intensive industrial park based on adaptive stepped demand response incentive mechanism[J]. Electric Power Systems Research, 2024, 233: 110504.
- [12] 周伟, 孙永辉, 王建喜, 等. 计及自适应阶梯型需求响应激励机制的园区综合能源系统优化调度[J]. 电网技术, 2023, 47(10): 4210-4218.
ZHOU Wei, SUN Yonghui, WANG Jianxi, et al. Optimal dispatch of park-level integrated energy system considering adaptive stepped demand response incentive mechanism[J]. Power System Technology, 2023, 47(10): 4210-4218(in Chinese).
- [13] LIN Jin, DONG Jun, DOU Xihao, et al. Psychological insights for incentive-based demand response incorporating battery energy storage systems: a two-loop Stackelberg game approach[J]. Energy, 2022, 239: 122192.

- [14] 肖白, 崔涵淇, 姜卓, 等. 基于有限理性用户选择行为的定制化电价套餐设计[J]. 电网技术, 2021, 45(3): 1050-1058.
XIAO Bai, CUI Hanqi, JIANG Zhuo, et al. Customized electricity price package design based on limited rational user selection behavior[J]. Power System Technology, 2021, 45(3): 1050-1058(in Chinese).
- [15] 柏颖, 刘晓峰, 陈雪颖. 面向需求响应日前市场投标的负荷聚合商决策动力学分析[J]. 电力系统自动化, 2022, 46(24): 95-103.
BAI Ying, LIU Xiaofeng, CHEN Xueying. Dynamics analysis on decision-making of load aggregator for day-ahead demand response market bidding[J]. Automation of Electric Power Systems, 2022, 46(24): 95-103(in Chinese).
- [16] 朱兰, 田泽清, 唐陇军, 等. 计及细节层次直接负荷控制的区域综合能源系统多时间尺度优化调度[J]. 电网技术, 2021, 45(7): 2763-2772.
ZHU Lan, TIAN Zeqing, TANG Longjun, et al. Multi-time-scale optimal dispatch of regional integrated energy system considering level of detail direct load control[J]. Power System Technology, 2021, 45(7): 2763-2772(in Chinese).
- [17] WEI Congying, WU Qiuwei, XU Jian, et al. Distributed scheduling of smart buildings to smooth power fluctuations considering load rebound[J]. Applied Energy, 2020, 276: 115396.
- [18] 祁兵, 郑顺林, 孙毅, 等. 考虑需求侧动态及耦合特性的激励型综合需求响应优化建模[J]. 中国电机工程学报, 2022, 42(5): 1783-1798.
QI Bing, ZHENG Shunlin, SUN Yi, et al. A model of incentive-based integrated demand response considering dynamic characteristics and multi-energy coupling effect of demand side[J]. Proceedings of the CSEE, 2022, 42(5): 1783-1798(in Chinese).
- [19] 马鹏, 齐军, 熊逸, 等. 考虑响应可靠性与经济性的需求侧精准调峰策略[J]. 电力系统自动化, 2025, 49(5): 176-186.
MA Peng, QI Jun, XIONG Yi, et al. Precise peak-shaving strategy at demand side considering response reliability and economy[J]. Automation of Electric Power Systems, 2025, 49(5): 176-186(in Chinese).
- [20] HAN Rushuai, HU Qinran, CUI Hantao, et al. An optimal bidding and scheduling method for load service entities considering demand response uncertainty[J]. Applied Energy, 2022, 328: 120167.
- [21] 葛晓琳, 胡文哲, 符杨, 等. 基于用户响应意愿度三维 Sigmoid 云模型的电动汽车优化调度策略[J]. 中国电机工程学报, 2024, 44(22): 8874-8883.
GE Xiaolin, HU Wenzhe, FU Yang, et al. Optimal scheduling strategy for electric vehicles based on user response willingness three dimensional Sigmoid cloud model[J]. Proceedings of the CSEE, 2024, 44(22): 8874-8883(in Chinese).
- [22] CHEN Lingyu, XU Jian, SUN Yuanzhang, et al. Bidding strategies of load aggregators for day-ahead market with multiple uncertainties[J]. IEEE Transactions on Power Systems, 2024, 39(2): 2786-2800.
- [23] KAUR K, KUMAR N, SINGH M. Coordinated power control of electric vehicles for grid frequency support: MILP-based hierarchical control design[J]. IEEE Transactions on Smart Grid, 2019, 10(3): 3364-3373.
- [24] 华昊辰, 翟家祥, 陈星莺, 等. 基于增强精英保留遗传算法的虚拟微网群动态划分及能量局域自治[J]. 中国电机工程学报, 2024, 44(12): 4652-4665.
HUA Haochen, ZHAI Jiaxiang, CHEN Xingying, et al. Dynamical partitioning and local energy autonomy of virtual microgrid groups based on strengthen elitist genetic algorithm[J]. Proceedings of the CSEE, 2024, 44(12): 4652-4665(in Chinese).
- [25] CHADUVULA H, DAS D. Analysis of microgrid configuration with optimal power injection from grid using point estimate method embedded fuzzy-particle swarm optimization[J]. Energy, 2023, 282: 128909.
- [26] HUA Haochen, QIN Yuchao, HAO Chuantong, et al. Optimal energy management strategies for energy Internet via deep reinforcement learning approach [J]. Applied Energy, 2019, 239: 598-609.
- [27] THIMMAPURAM P R, KIM J. Consumers' price elasticity of demand modeling with economic effects on electricity markets using an agent-based model[J]. IEEE Transactions on Smart Grid, 2013, 4(1): 390-397.
- [28] RUAN Guangchun, ZHONG Haiwang, SHAN Baoguo, et al. Constructing demand-side bidding curves based on a decoupled full-cycle process[J]. IEEE Transactions on Smart Grid, 2021, 12(1): 502-511.
- [29] MOREWEDGE C K, GIBLIN C E. Explanations of the endowment effect: an integrative review[J]. Trends in Cognitive Sciences, 2015, 19(6): 339-348.
- [30] 熊小萍, 杨辉, 蔡义明. 考虑用户禀赋效应和环保意识不确定性的微电网鲁棒优化调度方法[J]. 中国电机工程学报, 2023, 43(21): 8260-8269.
XIONG Xiaoping, YANG Hui, CAI Yiming. Robust optimal dispatch method of microgrid considering user

- endowment effect and environmental awareness uncertainty[J]. Proceedings of the CSEE, 2023, 43(21): 8260-8269(in Chinese).
- [31] LIU Di, CAO Junwei, LIU Mingshuang. Joint optimization of energy storage sharing and demand response in microgrid considering multiple uncertainties [J]. Energies, 2022, 15(9): 3067.
- [32] 华昊辰, 辛世禹, 陈星莺, 等. 基于“虚拟碳储存”的需求侧电-碳耦合交易机制[J]. 中国电机工程学报, 2024, 44(6): 2131-2143.
HUA Haochen, XIN Shiyu, CHEN Xingying, et al. Demand side electricity-carbon coupling trading mechanism based on “virtual carbon storage” [J]. Proceedings of the CSEE, 2024, 44(6): 2131-2143(in Chinese).
- [33] EGBOMWAN O E, LIU Shichao, CHAOUI H. Twin delayed deep deterministic policy gradient(TD3) based virtual inertia control for inverter-interfacing DGs in microgrids[J]. IEEE Systems Journal, 2023, 17(2): 2122-2132.
- [34] HUA Haochen, QIN Zhaoming, DONG Nanqing, et al. Data-driven dynamical control for bottom-up energy Internet system[J]. IEEE Transactions on Sustainable Energy, 2022, 13(1): 315-327.
- [35] BEHME A, DI TELLA P, SIDERIS A. On moments of integrals with respect to Markov additive processes and of Markov modulated generalized Ornstein-Uhlenbeck processes[J]. Stochastic Processes and their Applications, 2024, 174: 104382.
- [36] Pecan Street Inc. Dataport: the world's largest resource for residential energy use data[EB/OL]. [2026-05-25]. <https://dataport.pecanstreet.org/>.
- [37] ZHAO Xueyuan, GAO Weijun, QIAN Fanyue, et al. Electricity cost comparison of dynamic pricing model based on load forecasting in home energy management system[J]. Energy, 2021, 229: 120538.



华昊辰

在线出版日期: 2026-01-16。

收稿日期: 2025-03-28。

作者简介:

华昊辰(1988), 男, 博士, 教授, 主要研究方向为需求侧灵活资源调度、微网建模与控制、电力市场与电力经济, huahc16@tsinghua.org.cn。

(责任编辑 李泽荣)