

计及风电的发电商报价多智能体模型

黄飞虎, 李沛东, 彭舰, 董石磊, 赵红磊, 宋卫平, 李强

引用本文

黄飞虎, 李沛东, 彭舰, 董石磊, 赵红磊, 宋卫平, 李强. [计及风电的发电商报价多智能体模型](#)[J]. 计算机科学, 2024, 51(6A): 230600179-8.

HUANG Feihu, LI Peidong, PENG Jian, DONG Shilei, ZHAO Honglei, SONG Weiping, LI Qiang. [Multi-agent Based Bidding Strategy Model Considering Wind Power](#) [J]. Computer Science, 2024, 51(6A): 230600179-8.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[Camellia密码算法S盒的量子电路优化](#)

Quantum Circuit Optimization of Camellia Cryptographic Algorithm S-box
计算机科学, 2024, 51(6A): 230900051-6. <https://doi.org/10.11896/jsjcx.230900051>

[基于智能规划的多智能体强化学习算法](#)

Multi-agent Reinforcement Learning Algorithm Based on AI Planning
计算机科学, 2024, 51(5): 179-192. <https://doi.org/10.11896/jsjcx.230800099>

[基于观测重构的多智能体强化学习方法](#)

Multi-agent Reinforcement Learning Method Based on Observation Reconstruction
计算机科学, 2024, 51(4): 280-290. <https://doi.org/10.11896/jsjcx.230600055>

[基于隐空间扩散模型的差分隐私数据合成方法研究](#)

Differential Privacy Data Synthesis Method Based on Latent Diffusion Model
计算机科学, 2024, 51(3): 30-38. <https://doi.org/10.11896/jsjcx.230700177>

[稀疏异质多智能体环境下基于强化学习的课程学习框架](#)

Curriculum Learning Framework Based on Reinforcement Learning in Sparse Heterogeneous Multi-agent Environments
计算机科学, 2024, 51(1): 301-309. <https://doi.org/10.11896/jsjcx.230500146>

计及风电的发电商报价多智能体模型

黄飞虎^{1,2} 李沛东² 彭 舰² 董石磊¹ 赵红磊¹ 宋卫平¹ 李 强³

1 四川中电启明星信息技术有限公司 成都 610000

2 四川大学计算机学院 成都 610065

3 国网信息通信产业集团有限公司 北京 102211

(huangfh@scu.edu.cn)

摘 要 新型电力系统背景下,新能源发电商的报价问题一直是电力现货市场中的研究热点。相比传统能源,风电出力受外界不确定性因素的影响较大,给风力发电商求解最优报价带来了挑战。为此,基于多智能体强化学习算法 WoLF-PHC 构建了计及风电的发电商报价策略模型。模型中,考虑了风电、火电和水电 3 种能源参与的现货市场,每一个发电商抽象为一个智能体,且基于随机约束规划算法建模风电智能体的收益函数;对于智能体的报价策略模型,将 D3QN 与 WoLF-PHC 算法结合,使模型能够满足报价时智能体状态空间复杂的情况;此外,对于交互环境的建模,提出利用 DDPM 扩散模型生成风电出力数据,优化风电出清场景的仿真。最后,基于 3 节点的电力仿真系统开展模拟实验,实验结果表明,提出的风电收益函数建模、WoLF-PHC 改进、风电出力生成等技术是可行的,能有效解决风电参与竞价的现货市场报价问题,并且能够在较少的迭代次数后学习到较优的策略。

关键词: WoLF-PHC;多智能体强化学习;电力现货市场;竞价策略;扩散模型

中图分类号 TP389

Multi-agent Based Bidding Strategy Model Considering Wind Power

HUANG Feihu^{1,2}, LI Peidong², PENG Jian², DONG Shilei¹, ZHAO Honglei¹, SONG Weiping¹ and LI Qiang³

1 Aostar Information Technologies Co.,Ltd., Chengdu 610000, China

2 College of Computer Science, Sichuan University, Chengdu 610065, China

3 State Grid Information & Telecommunication Group Co.,Ltd., Beijing 102211, China

Abstract Under the background of new power system, the pricing problem of new energy generators has been a research hotspot in the electricity spot market. Compared with traditional energy, wind power output is subject to more uncertain factors, which poses a challenge to wind power generators in finding the optimal bid. To address this issue, this paper proposes a pricing strategy model for generators that takes into account of wind power based on the multi-agent reinforcement learning algorithm named WoLF-PHC. In the model, the spot market includes wind power, thermal power, and hydropower, and each generator is abstracted as an intelligent agent, and a stochastic constrained planning algorithm is used to model the profit function of the wind power agent. For the pricing strategy model of the agents, the D3QN algorithm is combined with the WoLF-PHC algorithm, which enables the model to handle complex state spaces when bidding. In addition, to model the interactive environment, a DDPM diffusion model is proposed to generate wind power output data and optimize the simulation of wind power clearing scenarios. In this paper, simulation experiments are carried out based on a 3-node power simulation system. Experimental results show that the proposed wind power profit function modeling, WoLF-PHC improvement, wind power output generation, and other techniques are feasible, which can effectively solve the bidding pricing problem of wind power in the spot market, and learn better strategy after fewer iterations.

Keywords WoLF-PHC, Multi-agent reinforcement learning, Electricity spot market, Bidding strategy, Diffusion model

1 引言

以来,我国首批 8 个现货试点(广东、蒙西、浙江、山西、山东、福建、四川、甘肃)已全部投入模拟试运行。同时,在新型电力系统背景下,新能源(风电和光伏)占有率不断增加^[1]。然而,

近年来,我国大力推进电力现货市场建设,至 2017 年

基金项目:四川省重点研发计划(2023YFG0112, 22ZYFG0034);四川省重点实验室开放课题(SCITLAB-20001);四川大学博士后交叉学科基金(10822041A2137);四川大学宜宾市合作项目(2020CDYB-30)

This work was supported by the Sichuan Science and Technology Program(22ZYFG0034), Intelligent Terminal Key Laboratory of Sichuan Province(SCITLAB-20001), Post doctoral Interdisciplinary Innovation Fund(10822041A2137) and Sichuan University and Yibin Cooperation Program(2020CDYB-30).

通信作者:彭舰(jianpeng@scu.edu.cn)

风电和光伏受外界条件如天气等因素的影响很大,出力的波动性强,给发电带来了很大的不确定性。针对新能源发电商参与电力现货市场竞价的场景,如何估计其不确定性是研究发电商报价技术需要应对的主要挑战^[2]。

目前,发电商报价技术可以总结为3类:(1)基于成本的分析方法,即通过经济学理论分析发电成本,然后加上利润作为最终报价^[3]。比如,文献[4]对发电商的内部成本与外部成本分别进行计算,然后分析各成本构成对发电竞价的影响并提出了基于成本分析的竞价策略。(2)基于博弈论的方法。该类方法根据电力交易过程构造博弈模型,然后求解模型均衡点^[5]。比如,文献[6]在假设竞价对手保持使用相同竞价策略进行报价的情况下讨论了不同交易量下的出价策略。(3)基于强化学习的方法,人们用智能体模拟发电商竞价过程,学习最佳的报价策略,这是目前的主流方法^[7-9]。比如,Kong等研究了电力现货日前市场中电力服务供应商的动态定价问题,将电力服务提供商建模为智能体学习合理的出价方案来应对未知用户响应行为的问题,并利用SARSA算法进行探索和求解^[10]。在一些场景中把发电商建模为智能体时会面临智能体状态空间过大的问题,因此一些工作使用深度强化学习算法来解决问题。比如,Gao等研究了日前市场中发电侧分时竞价场景下的竞价策略,为了解决智能体状态空间过大情况下难以求解和估计值函数大于真实值函数的问题,采用Double DQN算法作为求解方法^[11]。为了应对电力市场竞价环境的复杂多变和多智能体行为的相互影响,最近的工作倾向于使用多智能体强化学习方法来研究发电商的报价策略以辅助报价。Liu等研究了火力发电商在中长期市场中的辅助报价技术,将多个参与竞价的火力发电商建模为多智能体并利用MADDPG算法学习智能体的策略^[12]。Tang等提出了一种利用多智能体强化学习来模拟电力市场中发电商竞价行为的模型,使用WoLF-PHC算法更新智能体策略^[13]。

在解决新能源发电商带来的不确定性方面,研究者也做了一些探讨:出力场景生成和建立不确定性优化模型。例如,文献[14]采用生成网络生成以风电为代表的可再生能源出力场,并采用鲁棒性优化建模不确定性。文献[15]和[16]也基于生成对抗网络实现了指定类型的可再生能源场景生成。文献[2]通过Monte-Carlo抽样和聚类方法实现了风电出力的场景模拟和缩减,采用条件风险价值理论评估清洁能源和电价预测偏差风险,并得到最优竞价策略。

本文认为在发电商报价场景中,各发电商之间是不知道对方的报价思路的,每个发电商能够得到的激励反馈仅限于最后的出清结果。因此,对于多智能体建模所采用的算法应该考虑这种情况。而且,已有的基于强化学习的多智能体报价模型没有考虑风力发电商参与竞价的场景。此外,针对风电出力的不确定性,现有的场景生成方法在效果方面还可以进一步改进。本文针对现有工作的不足,提出了基于改进WoLF-PHC的多智能体报价模型。模型中,原始WoLF-PHC适用于建模竞争对手策略位置的情况,但其缺点是不足以处理状态空间集很大的情况,为此本文引入了Double-Q网络增强对状态空间的建模能力,实验结果表明这种改进也缩短了策略求解时间。同时,本文提出了使用扩散模型DDPM(Denoising Diffusion Probabilistic Models)生成风电出力的

场景^[17]。该模型在图像生成领域比生成对抗网络有更好的效果。从实验效果来看,利用该模型生成的风电出力数据也具有更好的质量。

2 多智能体强化学习

多智能体强化学习是强化学习的一个重要分支^[18]。多个智能体根据自身所处的状态和各自的策略选择动作与环境进行交互从而达到下一个状态和获取奖励。智能体之间的关系包括协作任务和竞争任务,协作任务中智能体之间共享一个目标函数,通过相互交流和与环境交互实现策略的学习;竞争任务中各智能体之间目标函数独立且相互竞争,智能体通过观察其他智能体的策略和与环境交互来更新自己的策略。

WoLF-PHC算法属于竞争任务中的一种多智能体强化学习算法^[19],它是WoLF规则(Win or Learn Fast,简称WoLF)与PHC算法(policy hill-climbing,简称PHC)的结合。WoLF规则,即当智能体做的比期望值好的时候,小心缓慢地调整参数,当智能体做的比期望值差的时候,加快步伐调整参数。PHC算法的思路是增大能够得到最大累积期望的动作用的选取概率。给定状态空间 S ,动作空间 A ,智能体 i 的 Q 函数表示为 $Q_i(s, a)$,策略表示为 $\pi_i(s, a)$, $C(s)$ 表示状态 s 出现的次数。WoLF-PHC算法的主要流程如算法1所示。

算法1 WoLF-PHC

1. 初始化每个智能体的参数: $\pi_i(s, a) = \frac{1}{|A_i|}$, $Q_i(s, a) = 0$, 平均策略
 $\bar{\pi}(s, a) = \frac{1}{|A_i|}$, $C(s) = 0$, $\delta_i > \delta_w$;
2. for t in $(1, T)$ do:
3. 对于每个智能体 i 根据当前状态 s_t 采用探索-利用策略选择动作 a_t ;
4. 计算奖励 r 和下一个状态 s_{t+1} ;
5. 更新 $Q_i(s_t, a_t)$:
 $Q_i(s_t, a_t) \leftarrow Q_i(s_t, a_t) + \alpha[\gamma_i + \max_{a'} Q_i(s', a') - Q_i(s_t, a_t)]$
6. 对每个 $a_j \in A_i$,更新平均策略 $\bar{\pi}_i(s_t, a_j)$:
 $C(s) = C(s) + 1$
 $\bar{\pi}_i(s_t, a_j) = \bar{\pi}_i(s_t, a_j) + \frac{1}{C(s)}[\pi_i(s_t, a_j) - \bar{\pi}_i(s_t, a_j)]$
7. 根据 $Q_i(s_t, a_j)$,对每个 $a_j \in A_i$,更新 $\pi_i(s_t, a_j)$:
 $\pi_i(s_t, a_j) \leftarrow \pi_i(s_t, a_j) + \Delta_{sa_j}$
 其中, $\Delta_{sa_j} = \begin{cases} -\delta_{sa_j}, & \text{if } a_j \neq \arg\max_{a'} Q(s_t, a') \\ \sum_{a' \neq a_j} \delta_{sa'} & \text{else} \end{cases}$
 $\delta_{sa_j} = \min(\pi_i(s_t, a_j), \frac{\delta}{|A_i| - 1})$
 $\delta = \begin{cases} \delta_w, & \text{if } \sum_{a_j \in A_i} \pi_i(s_t, a_j) Q_i(s_t, a_j) > \sum_{a_j \in A_i} \bar{\pi}_i(s_t, a_j) Q_i(s_t, a_j) \\ \delta_i, & \text{else} \end{cases}$
8. end for

在WoLF-PHC算法中,使用一个可变的学習速率 δ 来实现WoLF效果,当策略效果较差时使用 δ_i ,策略效果较好时使用 δ_w ,并且满足 $\delta_i > \delta_w$ 。从执行步骤中可以看出,各智能体不需要知道竞争对手的策略和反馈。这一特点非常适合电力市场中的报价场景,各发电厂商仅需根据出清的结果更新自己的策略即可。基于这一特点,每个智能体只需保存自己的动作就可以完成策略学习。相比于Minimax-Q, Nash Q-Learning,

FFQ等算法,各智能体可用更少的空间去记录Q值,从而极大地减小了存储空间。

3 本文模型

3.1 相关定义

定义1(智能体) 用集合 $G = \{g_1, g_2, \dots, g_n\}$ 表示,在电力现货日前市场中,发电商 i 为一个智能体(即 g_i),各智能体之间具有竞争关系。

定义2(最优报价策略) 对于智能体 g_i , π 为其报价策略,则最优报价策略 π^* 需满足:

$$\begin{cases} \operatorname{argmax}_{\pi^*} f(\pi) \\ \text{s. t. } g_i(\pi) \leq 0, i=1, 2, \dots, p \\ h_j = 0, j=1, 2, \dots, q \end{cases} \quad (1)$$

其中,函数 f 是收益函数,期望给出的报价策略 π 用于实现最大化利益;用函数 g 和 h 建模机组运行特性与市场边界条件。

定义2中,机组运行特性包括机组优先电量约束、机组出力上下限约束、机组爬坡约束等;市场边界条件包括线路潮流

约束、电量空间边界等。在电力现货中,发电商报价的目标就是根据自身的发电成本、机组出力等条件、市场边界条件以及竞争对手的情况给出最优的报价策略 π^* ,实现利润最大化,这是带约束条件的优化问题。

3.2 模型框架

在电力现货市场的交易可以分为两个阶段:首先,各发电商根据独立调度机构在交易日前一天发布的负荷需求量(交易日)制定报价策略并提交给市场运营机构;然后,市场运营机构根据各发电商的报价情况,在满足电力系统稳定运行的前提下进行统一出清,并将结果反馈给各发电商^[1]。在强化学习中,交互环境即是智能体收益的反馈方。本文中,各发电厂厂商作为智能体,收益就是每次报价的中标结果。因此,交互环境就是电力现货市场工作流程,总体建模思路如图1所示。模型分为以下核心模块:(1)发电商建模,根据各能源特点进行建模;(2)报价策略模型,为各发电商提供报价;(3)交互环境建模,模拟市场出清过程;(4)风电场景生成,在交互环境中模拟风电出力的不确定性。下面将详细介绍各模块。

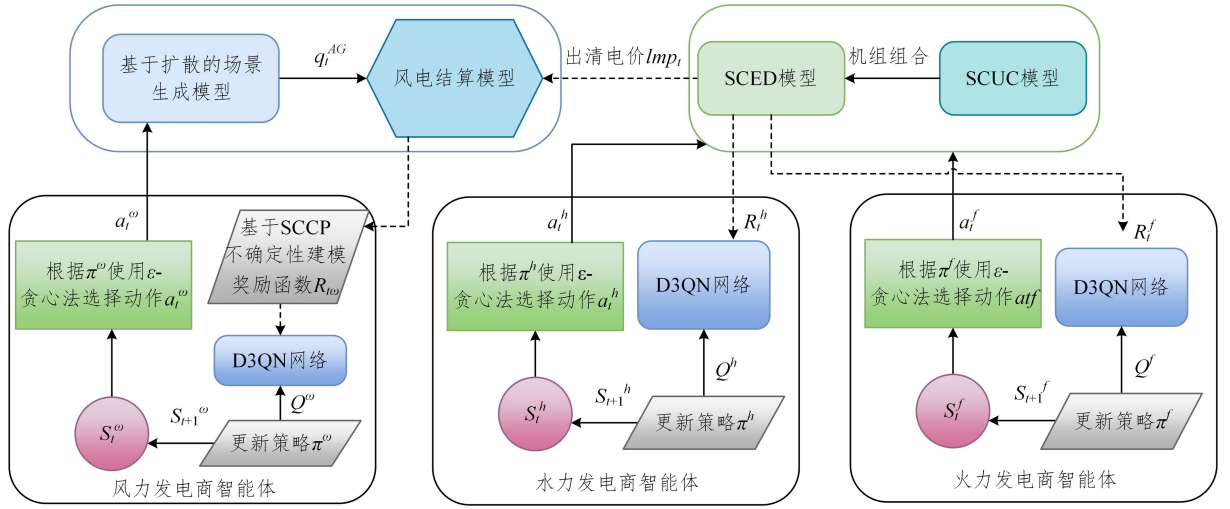


图1 模型框架

Fig. 1 Framework of the proposed model

3.3 发电商建模

本文主要考虑有风电接入的,由风电、水电和火电3种能源构成的市场环境,且不同的发电商建模为一个智能体。在强化学习中,每个智能体具有动作空间、状态空间和收益函数,本小节将作详细阐述。

(1) 动作空间

在现货市场中,发电商的目标是给出最优报价策略,因此,可将其每个动作 $a_i(x)$ 建模为一组报价曲线,由于报价与出力及价格相关,因此将其定义为一个分段函数:

$$a_i(x) = \begin{cases} p_{i,1}, & k_1 \leq x < k_2 \\ p_{i,2}, & k_2 \leq x < k_3 \\ \dots & \dots \\ p_{i,d}, & k_{d-1} \leq x < k_d \end{cases} \quad (2)$$

其中, x 表示出力,给定发电商的出力区间 $[P_{\min}, P_{\max}]$,分为 d 段,即 $[k_1, k_2)$ 到 $[k_{d-1}, k_d)$; $p_{i,\cdot}$ 表示对应区间的价格,按照报价规则(随着出力增加,价格递增),则 $p_{i,1} \leq p_{i,2} < p_{i,d}$ 。综上,动作空间 A 建模为一组 $a_i(x)$ 的集合, $a_i(x)$ 简写为 a_i ,即 $A = \{a_1, a_2, \dots, a_n\}$ 。

(2) 状态空间

状态空间 $S = \{s_1, s_2, \dots, s_n\}$,定义为发电机的实际出力,即中标电量。给定发电商的出力区间 $[P_{\min}, P_{\max}]$,平均划分为 n 个区间,每个区间 $[P_i, P_{i+1})$ 对应状态 s_i 。由于水电的出力不仅是发电机的最大出力决定,还与实际水量相关,所以对于建模水电发电商的智能体,最大出力区间 $[P_i, P_{i+1})$ 需要根据实际的水量极限和发电机最大出力调整,即 $P_{i+1} = \max(Q, P_{\max})$,其中 Q 表示水量极限。

(3) 收益函数

对于每个智能体,其策略 π 定义为在状态 s 时选择动作 a 的概率分布。策略 π 需要根据收益反馈不断地学习更新。下面将分别介绍3种能源的发电商的收益函数或目标函数。各智能体的策略学习目标是最大化收益函数。

火力发电商的收益函数为:

$$B^h = \sum_{t=1}^T (q_t^h P_t^h - C_t^h) \quad (3)$$

其中, q_t^h , P_t^h 和 C_t^h 分别表示 t 时间的中标电量、电价和成本。成本的计算公式为:

$$C_t^h = a_1 (q_t^h)^2 + a_2 q_t^h + a_0 \quad (4)$$

其中, a_0, a_1, a_2 是成本系数。

水力发电商的收益函数为:

$$B^f = \sum_{t=1}^T (q_t^s P_t^s - C_t^s) \quad (5)$$

$$C_t^s = a_1 (q_t^s)^2 + a_2 q_t^s + a_0 \quad (6)$$

对于风力发电商, 由于其出力受天气影响很大, 具有不确定性。同时, 现货市场的出清价格也有一定的不确定性。为了建模风电出力和市场反馈的不确定性, 本文采用随机机会约束规划 (Stochastic Chance Constrained Programming, 简称 SCCP)^[20], 将风力发电商的收益函数设计为:

$$\begin{cases} B^f = \sum_{t=1}^T (q_t^s P_t^s - C_t^s) \\ \text{s. t. } Pr(B^f \geq B^*) \geq \alpha \end{cases} \quad (7)$$

其中, B^* 为期望收益, 计算公式为:

$$B^* = \sum_{t=1}^T (q_t^* P_t^f - C_t^f) \quad (8)$$

$$C_t^f = a_0 q_t^f \quad (9)$$

SCCP 适用于约束条件中含有随机变量, 并需要考虑到所做决策在不利情况发生下不满足约束条件的可能性, 且必须在预测到随机变量的取值之前做出决策的情况。因此, 式(7)的设计思路是当期期望收益为 B^* 时, 概率 $Pr(B^f \geq B^*)$ 不小于 α 。 q_t^* 是对未来中标电量的估计, 本文基于历史真实发电量进行采样求均值得到。

此外, 由于风电的并网容量比水电和火电少, 而且不确定因素大。在建模时, 本文设置其不参与价格竞争, 仅作为市场价格接受者^[2]。具体为, 若其报价 $P_t^f < lm p_t$ 市场出清的节点边际价格时, 中标成功, 中标价格为 P_t^f ; 否则, 中标失败。表示为:

$$P_t^f \begin{cases} P_t^f, & P_t^f \leq lm p_t \\ 0, & P_t^f > lm p_t \end{cases} \quad (10)$$

3.4 报价策略模型

本文基于 WoLF-PHC 算法构建各智能体的报价策略模型。在原始 WoLF-PHC 算法中, 采用 Q-Learning 更新动作价值函数, 其缺点是无法有效应对智能体状态空间复杂的情况, 并且求解速度不够快。因此, 本文采用 Dueling Double DQN (D3QN) 网络来计算各个状态对应动作的 Q 函数值, 如图 2 所示。

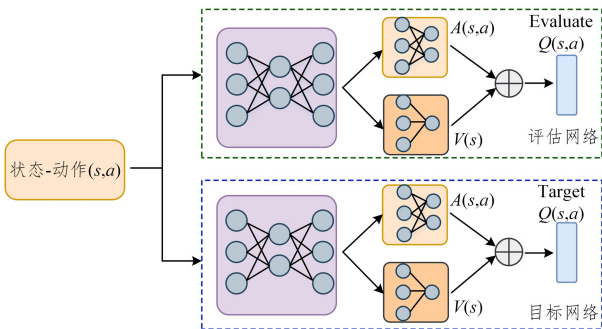


图 2 D3QN 网络

Fig. 2 D3QN network

基于算法 1 的流程, 引入 D3QN 的 WoLF-PHC 算法过程如算法 2 所示。在第 4-16 步中, 更新 Q 函数值的方式改进为使用 Dueling Double DQN (D3QN) 网络 (见图 2)。D3QN 采用经验回放集合来训练网络参数, 在更新网络参数的阶段, 将四元组 $(s_t, a_t, R_{t+1}, s_{t+1})$ 放入经验回放集合中。

进行网络参数的更新时, 对经验回放集合进行 M 次采样, 然后计算每个样本的目标 Q 函数值。为了避免 Q 函数被高估, D3QN 在计算非终止状态的目标 Q 函数值时采用了 Double DQN 算法, 即解耦选择目标 Q 函数值动作和计算目标 Q 函数值这两个步骤, 如算法第 11 步所示。最后将目标 Q 函数值和当前值网络计算出的 Q 函数值之差作为损失函数, 使用梯度反向传播更新网络参数。目标值网络的参数通过定期赋值的方式同步更新。

算法 2 改进的 WoLF-PHC

输入: 各个智能体的状态空间 S , 动作空间 A , Q 函数值学习率 α , 策略学习率 δ_w 和 δ_l , 奖励衰减因子 γ , 经验回放集合采样数 M , 经验回放集合大小 MS , 状态计数器 $C(s)$, 训练的最大迭代轮数 T
输出: 各个智能体的策略函数 π

1. 对每个智能体进行初始化: $\pi(s, a) = \bar{\pi}(s, a) \frac{1}{|A|}$, $C(s) = 0$, 随机初始化当前值网络参数 $\omega = (\omega^a, \omega^b)$, 基于 ω 初始化当前值网络中所有状态下对应各个动作的 Q 函数值, 清空经验回放集合, 其中 Q 函数值的计算方式为:

$$q_\omega(s_t, a_t) = V_\omega^*(s_t) + A_{\omega^a}(s_t, a_t) - \frac{1}{|A|} \sum_{a \in A} A_{\omega^a}(s_t, a)$$

$\omega' \leftarrow \omega$, 基于 ω' 初始化目标值网络中所有状态下对应各个动作的 Q 函数值

2. for 1-T 轮迭代对每一个智能体 do

3. while 当前状态 s_t 不是终止状态 do

4. 根据当前状态 s_t 和当前策略 π 使用 ϵ -贪心法选择动作 a_t

5. 与环境交互得到下一个状态 s_{t+1} 和延时奖励 R_{t+1}

6. if 经验回放集合已满 then

7. 根据丢弃策略丢弃已有的四元组

8. end if

9. 将四元组 $(s_t, a_t, R_{t+1}, s_{t+1})$ 存入经验回放集合中

10. for 1-M 轮采样 do

11. 对于第 i 个样本 $(s_t, a_t, R_{t+1}, s_{t+1})_i$, 计算当前值网络的目标 Q 函数值 y_i :

$$y_i = \begin{cases} R_{t+1}, & s_{t+1} \text{ 是终止状态} \\ R_{t+1} + \gamma q_{\omega'}(s_{t+1}, \arg\max_a q_\omega(s_{t+1}, a)), & \text{否则} \end{cases}$$

12. 计算损失函数: $Loss = y_i - q_\omega(s_t, a_t)$, 并使用梯度反向传播更新当前值网络参数 ω

13. if 参数 ω 已经经过 N 次更新 then

14. $\omega' \leftarrow \omega$

15. end if

16. end for

17. 更新平均策略函数值:

$$C(s_t) = C(s_t) + 1$$

$$\bar{\pi}(s_t, a_t) = \bar{\pi}(s_t, a_t) + \frac{1}{C(s_t)} [\pi(s_t, a_t) - \bar{\pi}(s_t, a_t)]$$

18. 使用当前值网络输出状态 s_t 和 s_{t+1} 下的 Q 函数, 更新策略函数值:

$$\pi(s_t, a_t) = \pi(s_t, a_t) + \Delta_{(s_t, a_t)}$$

其中

$$\Delta_{(s_t, a_t)} = \begin{cases} -\delta_{(s_t, a_t)}, & a_t \neq \arg\max_a Q(s_{t+1}, a) \\ \sum_{a \neq a_t} \delta_{(s_t, a)}, & \text{否则} \end{cases}$$

$$\delta_{(s_t, a_t)} = \min(\pi(s_t, a_t), \frac{\delta}{|A| - 1})$$

$$\delta = \begin{cases} \delta_w, & \text{if } \sum_{a \in A} \pi(s, a) Q(s, a) > \sum_{a \in A} \bar{\pi}(s, a) Q(s, a) \\ \delta_l, & \text{否则} \end{cases}$$

19. $s_t \leftarrow s_{t+1}$
 20. end while
 21. end for

3.5 交互环境建模

图 1 中,火电、水电和风电 3 种能源发电商被分别建模为一个智能体。火力发电商智能体和水利发电商智能体将各自的动作(报价)提交给负责市场出清的环境子模块,该子模块由 SCUC(Security Constrained Unit Commitment)模型负责根据发电商的报价与机组特性、电力系统约束等条件计算出能够在保证电力系统安全运行前提下发电总成本最小的机组启停组合。然后将机组启停组合作为 SCED(Security Constrained Economic Dispatch)模型的输入,SCED 模型通过优化发电机的实际出力与负荷需求的偏移量来保证发电商的利润从而计算得到发电商的中标电价^[21]。

对于风力发电商智能体,本文设置其不参与价格竞争,仅作为现货市场的价格接受者。同时,为了建模其出力的不确定性。在交互环境中,通过风力发电机的历史出力数据来生成符合风力发电机的出力特征,并将出力数据用于风电结算模型计算风力发电商智能体的中标电量与中标电价。

3.6 场景生成模型

场景生成模型负责模拟出风力发电机的发电量,风力发电机由于受天气、季节等外在因素的影响,发电量通常会存在剧烈波动。为了解决模拟风力发电机真实出力场景的问题,目前的主流方法是借助深度学习方法来解决,比如使用 GAN 模型,通过风力发电机的历史发电量数据合成风力发电机的出力场景^[22-24]。

近年来扩散模型成为了深度生成模型领域的新热点,在图像生成任务中取得了比 GAN 模型更好的效果^[25]。因此,本文认为扩散模型用于风电出力场景生成的效果会优于基于 GAN 的模型,并提出基于 DDPM(Denoising Diffusion Probabilistic Models)^[17]构建风电出力场景生成模型,实验结果也实证了这一观点。

模型的工作流程如算法 3 所示,模型由正向扩散(第 2-10 步)和逆扩散(第 11-19 步)两个步骤组成。将历史风电出力数据作为模型输入,通过若干步加入已知噪声使数据服从各向同性的高斯分布。再通过逆扩散过程从加噪后的数据逐步还原出真实数据。该模型使用一个神经网络在逆扩散过程中预测每一步添加的噪声,而将正向扩散使用的已知噪声作为标签。网络输入为加噪后的数据和加噪步数,输出为预测的噪声。训练好该神经网络的参数后就可以将其用于还原真实数据,达到生成符合真实数据统计特点的数据的效果。

算法 3 基于扩散的风电出力场景生成模型

输入:风力发电机的历史出力数据 X,最大扩散次数 T,加噪权重初始值 β_0 和它的变化范围 $[\beta_1, \beta_T]$

输出:生成的风力发电机出力数据

1. 对网络参数 θ 进行随机初始化
2. while 网络没有收敛 do
3. 在真实历史数据分布 $q(X)$ 中采样出训练数据 x_0
4. 随机生成一个扩散次数 $t^*, t^* \in [1, T]$
5. 从标准正态分布中采样一个噪声 \tilde{n}
6. 进行 t^* 步的扩散:在第 $t(t \in [1, t^*])$ 步加噪时:

$$\beta_t \leftarrow \beta_{t-1} + \frac{\beta_T - \beta_{t-1}}{t^*}$$

数据 x_t 由上一步加噪后的数据 x_{t-1} 经过下式得到:

$$x_t = \sqrt{1 - \beta_t} x_{t-1} + \sqrt{\beta_t} n_{t-1}$$

7. 通过一次计算从初始数据 x_0 得到加噪 t 步后的数据 x_t :

令 $\alpha_t = 1 - \beta_t, \bar{\alpha}_t = \alpha_1 \times \alpha_2 \times \dots \times \alpha_t$, 则有

$$\begin{aligned} x_t &= \sqrt{\alpha_t} x_{t-1} + \sqrt{1 - \alpha_t} n_{t-1} \\ &= \sqrt{\alpha_t} \times (\sqrt{\alpha_{t-1}} x_{t-2} + \sqrt{1 - \alpha_{t-1}} n_{t-2}) + \sqrt{1 - \alpha_{t-1}} n_{t-1} \\ &= \dots \\ &= \sqrt{\alpha_t} x_0 + \sqrt{1 - \alpha_t} \tilde{n} \end{aligned}$$

8. 向网络输入加噪后的数据 x_t 和时间步 t 得到一个预测噪声 \tilde{n}_θ
9. 计算损失函数: $Loss = \|\tilde{n} - \tilde{n}_\theta(x_t, t)\|^2$, 并使用梯度反向传播更新网络参数 θ
10. end while
11. 逆扩散过程生成风力发电机出力数据:从标准正态分布中采样一个随机噪声 x_T 作为扩散 T 次后的数据
12. for $T-1$ 步逆扩散 do
13. if $t > 1$ then
14. 从标准正态分布中采样一个随机噪声 n_t
15. else
16. $n_t = 0$
17. end if
18. 从数据 s_t 逆扩散得到数据 s_{t-1} :

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \alpha_t}} n_\theta(x_t, t) \right) + \sigma_t n_t$$
19. end for

4 实验

4.1 实验设置

实验的仿真电力系统在 3 节点电力系统^[26]基础上进行了调整,加入了水力发电商与风力发电商。它的拓扑结构如图 3 所示。各个发电商拥有一个发电机组,各发电机组的参数如表 1 所列。

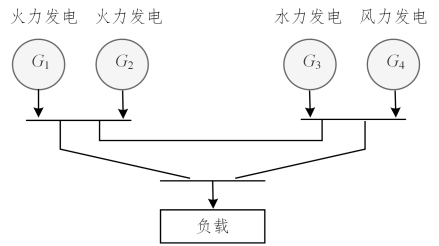


图 3 电力系统示意图

Fig. 3 Power system diagram

表 1 发电机组参数

Table 1 Generator set parameters

属性	G1	G2	G3	G4
最小出力/MW	5	5	5	0
最大出力/MW	75	75	73	70
容量/MW·h	2000	2000	1500	1300
启停成本/万元	80	72	68	10
成本参数 $a_2/(t/(MW \cdot h)^2)$	0.000068	0.000096	0.000018	—
成本参数 $a_1/(t/(MW \cdot h))$	0.16	0.22	0.26	—
成本参数 $a_0/(t/h)$	58	25	16	9
爬坡能力/(MW/h)	50	40	40	40

实验中,模拟数据分别来源于 PJM 网站(负荷数据)、福堂水电站运行数据(水电出力数据)和百度 KDD 比赛(风电出力数据)。负荷数据是从 PJM 网站下载的特定区域 2021 年 6 月 1 日-9 月 30 的历史负荷数据;同理,水电和风电的

出力数据也截取了 4 个月的数据。预处理时,把这些数据划分为若干个连续 15 天的数据集用于模型的训练与验证,即模型模拟电力市场运行 15 天。另外,统计了水力发电机的历史每日最大出力数据用于确定水力发电机的模拟水量极限。风力发电机的历史出力数据用于生成风力发电机的出力场景和不确定性建模。

在 D3QN 中,当前值网络和目标值网络都使用双层的全连接网络,隐藏层维度设为 64;激活函数使用 ReLU。经验回放集的大小为 2000,目标值网络的参数每 30 轮迭代更新一次,学习率设置为 0.1。在学习策略阶段,两个学习率 δ_w 和 δ_i 分别设置为 0.01 和 0.02,探索率 ϵ 设置为 0.1,奖励衰减因子为 0.95。风电场景生成模型使用了一个全连接网络,它拥有 3 个隐藏层,隐藏层维度设为 128,激活函数使用 ReLU。

4.2 有效性验证

该部分实验验证了基于扩散的风电场景出力模型和基于随机机会约束规划理论的风电不确定性建模两个子模块的有效性。

(1) 风电场景生成模型

将该模型与 3 个主流的风电场景生成模型(随机规划采样(SP)^[27]、生成对抗网络(GAN)^[16]和对抗变分贝叶斯模型(AVB)^[14]进行了对比。实验统计了生成数据的 Berier 分数和关键统计数据。Berier 分数主要用于衡量每一时刻生成数据 y 与真实数据 Y 的偏差,如式(11)所示,得分越低则效果越好,计算公式如下:

$$Berier(y) = \frac{\sum_{t=1}^T (E(Y_t) - y_t)^2}{T} \quad (11)$$

对于 SP 采样方法,直接在训练数据上进行采样生成风力发电机出力数据;对于 GAN、SeqGAN、AVB 和本文提出的基于扩散的风电出力场景生成模型分别进行了 2000 轮的训练,经过实验,这些模型均能够在训练周期之内收敛。然后根据上述指标比较了上述方法生成的风力发电机出力数据的质量。

各模型的生成数据与真实数据的 Berier 分数和各项关键统计数据的计算结果如表 2 所列,其中每一项的最佳值都加粗标出。

表 2 生成模型效果对比

统计项	真实数据	SP	GAN	AVB	本文模型
Berier	—	598.32	441.42	450.34	283.05
最大值/MW	55.0	54.4	52.2	50.2	55.0
85 百分位值/MW	50.5	52.7	34.2	44.4	48.6
中位数/MW	38.0	48.0	29.4	23.5	35.2
15 百分位值/MW	15.4	31.6	7.2	8.3	13.3
最小值/MW	7.3	5.0	5.0	5.0	5.5
平均值/MW	35.1	43.9	27.4	25.9	32.9
标准差/MW	13.7	11.7	12.4	15.0	14.7

从实验结果可以看出,本文模型实现了最佳的 Berier 分数,并且几乎每一项关键统计数据都是最贴近真实数据的。综上所述,实验结果的各项指标对比都验证了本文提出的基于扩散的风电出力场景生成模型的有效性与优势。相较于其他对比方法,SP 采样因为忽略了时序数据之间的关系和数据特征而很难生成效果较好的数据,SeqGAN 方法与 AVB 方法作为经典的生成模型 GAN 方法的改进模型,效果相较于 GAN 方法有了一定的提升,其中 SeqGAN 模型的效果是对比算法中最好的,这可能是由于它更多地捕捉风电出力历史数据的时间相关性来进行网络训练,然而训练时间较长和难以收敛是它仍然存在的问题。

(2) 不确定性建模

接下来本文验证了基于随机机会约束规划理论建模风电不确定性的有效性。不使用不确定性建模的记为 MARL,使用不确定建模的记为 MARL-SCCP。它们的对比结果如图 4 所示。比较指标有:(1)风力发电商 G4 的收益,它直接反映了风力发电商智能体在使用学习到的报价策略下能够获取的利润;(2)收益波动(Revenue Fluctuation, RF),通过统计风力发电商智能体 G4 每轮收益的方差然后进行归一化得到,以反映风力发电商收益的稳定性,数值越大代表波动越大,即风力发电商面临的不确定性越大;(3)中标率,该指标体现了风力发电商在一轮迭代中标天数占总天数的比例,数值越大代表中标天数越多。从实验结果来看,使用基于随机机会约束规划理论建模风电不确定性能够提高风力发电商的中标率和收益,并且在一定程度上降低了收益波动。

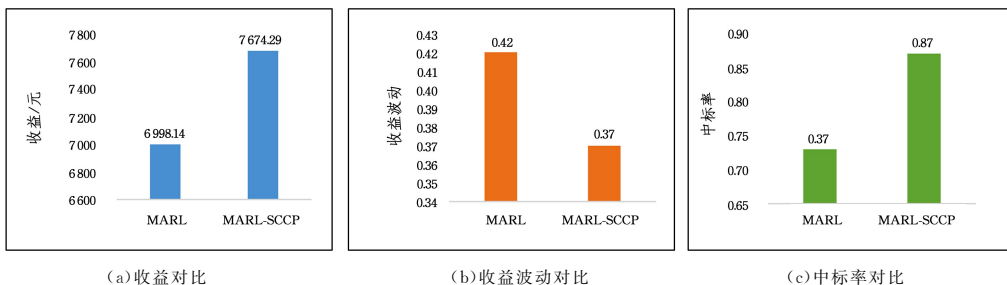


图 4 不确定性建模的有效性验证

Fig. 4 Validity verification of uncertainty modeling

4.3 策略模型对比

本节通过将求解算法替换为其他主流算法来对比测试结合 D3QN 的 WoLF-PHC 算法的性能。对比的算法有 Q-learning^[28], SARSA^[29], DQN^[30], Double DQN^[11] 和 WoLF-PHC^[13]。

实验对比了各算法的平均收敛时间(Average Conver-

gence Time, 简称 ACT)、平均总收益(Average Total Revenue, 简称 ATR)和风力发电商的收益波动(Average Revenue Fluctuation, 简称 ARF),结果如表 3 所列。在风力发电商参与下的电力现货日前市场竞价场景中,本文模型的收敛时间仍然是最快的,并且本文模型能够实现最大的发电商总收益。对于风力发电商来说,其收益波动也是最小的。对于单个发

电商智能体的平均收益,本文模型实现了几乎所有发电商中的最高收益。实验结果充分证明了采用改进的 WoLF-PHC 算法框架作为考虑不确定性的新能源发电商报价策略模型的求解算法的优势。

表 3 策略模型对比

Table 3 Comparison of strategy models

算法	ACT	ATR/元	ARF
Q-learning	167(±40.41)	88103.81	1.05
SARSA	>200	86745.62	1.03
DQN	76(±31.10)	87676.25	1.11
Double-DQN	78(±41.31)	86712.98	1.08
WoLF-PHC	160(±22.91)	88451.78	1.02
本文模型	38(±14.14)	88881.63	0.96

Q-learning 和 SARSA 在收敛速度上比其他模型慢,其主要原因是对于 Q 值函数的处理没有采用神经网络,一方面不能满足复杂的状态空间,另一方面,采用查询更新的方式速度较慢。DQN 和 Double DQN 则采用神经网络拟合 Q 值函数,收敛速度很快,拟合效果也较好,但这两种网络相比 D3QN 的表示能力还是略差。

结束语 本文研究了风力发电参与的电力现货的报价策略模型设计问题,并提出了基于多智能体强化学习算法 WoLF-PHC 的报价策略模型。模型采用随机机会约束规划建设风电出力 and 市场反馈的不确定性;同时,引入 D3QN 改进了 WoLF-PHC 中对于 Q 值函数的学习,增强模型应对智能体状态空间复杂的情况。此外,在环境建模中,提出采用 DDPM 构建风电出力场景生成模型,优化风电的出清模拟。最后,通过模拟实验验证了模型的有效性。当然,在实验部分,由于实验条件和公开信息有限,本文实验均在搭建的仿真环境下进行,存在缺乏考虑发生输电阻塞场景等问题。在未来的工作中,需要进一步优化仿真环境以反映真实的电力系统运行状况。

参考文献

- [1] WEN R, TAN L. Short-term Power Forecasting for Photovoltaic Generation Based on HS-ESN [J]. *Computer Science*, 2017, 44(6): 226-231, 265.
- [2] LU C, JIANG T, DENG H, et al. Bidding Strategy and Profit Distribution of Power Generation Company with Clean Energy in Spot Market Based on Cooperative Game Theory [J]. *Electric Power Construction*, 2020, 41(12): 9.
- [3] NING B. Analysis on generating cost of thermal power plant and its pricing strategy [D]. North China Electric Power University, 2009.
- [4] ZHANG L. Cost Analysis and Bidding Strategy Research of Power Plants under Power Market Environment [D]. North China Electric Power University, 2010.
- [5] WU J, WANG J, YAN Y. Agent-based dynamic simulation of an electricity market with multilateral bidding [J]. *International Journal of Modeling, Simulation, and Scientific Computing*, 2019, 10(3): 15-27.
- [6] YANG M. Research on bidding strategy of electricity market based on current trading mode [D]. Beijing University of Posts and Telecommunications, 2019.
- [7] WANG J, WU J, KONG X. Multi-agent simulation for strategic bidding in electricity markets using reinforcement learning [J]. *Power and Energy Systems*, 2020, 9(3): 1051-1065.
- [8] CHEN L, GUO T, LIU Y T, et al. Survey of Multi-Agent Strategy Based on Reinforcement Learning [C] // 2020 Chinese Control and Decision Conference (CCDC). 2020: 604-609.
- [9] DONG H, DONG H, DING Z, et al. Deep Reinforcement Learning [M]. Singapore: Springer, 2020.
- [10] KONG D, KONG X, XIAO J, et al. Dynamic pricing of demand response based on elasticity transfer and reinforcement learning [C] // 2019 22nd International Conference on Electrical Machines and Systems (ICEMS). IEEE, 2019: 1-5.
- [11] GAO Y, LI J, CAO R, et al. Simulation of Generators' Bidding Behavior Based on Multi-agent Double DQN [J]. *Power System Technology*, 2020, 44(11): 4175-4183.
- [12] LIU D, GAO Y, WANG W, et al. Research on bidding strategy of thermal power companies in electricity market based on Multi-Agent Deep Deterministic Policy Gradient [J]. *IEEE access*, 2021, 9: 81750-81764.
- [13] TANG C, ZHANG L, LIU F, et al. Research on Pricing Mechanism of Electricity Spot Market Based on Multi-agent Reinforcement Learning (Part I): Bi-level Optimization Model for Generators Under Different Pricing Mechanisms [J]. *Proceedings of the CSEE*, 2021, 41(2): 536-552.
- [14] JIANG T, WANG X, JIANG C, et al. Optimal Hybrid Stochastic Robust Bidding Strategy of Wind and Hydraulic Pumped Storage Jointly Participating in Day-ahead and Real-time Market Using Data-driven Method [J]. *Power System Technology*, 2022, 46(2): 481-495.
- [15] DONG X, SUN Y, PU T. Day-ahead scenario generation of renewable energy based on conditional GAN [J]. *Proceedings of the CSEE*, 2020, 40(17): 5527-5536.
- [16] CHEN Y, WANG Y, KIRSCHEN D S, et al. Model-free renewable scenario generation using generative adversarial networks [J]. *IEEE Transactions on Power Systems*, 2017, 33(99): 3265-3275.
- [17] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models [J]. *Advances in Neural Information Processing Systems*, 2020, 33: 6840-6851.
- [18] YUAN W, LUO J, LU L, et al. Methods in Adversarial Intelligent Game: A Holistic Comparative Analysis from Perspective of Game Theory and Reinforcement Learning [J]. *Computer Science*, 2022, 49(8): 191-204.
- [19] QIN Z, LI N, LIU X, et al. Overview of Research on Model-free Reinforcement Learning [J]. *Computer Science*, 2021, 48(3): 180-187.
- [20] ZHANG X, GAO W, ZHONG J. Decentralized Economic Dispatching of Multi-Micro Grid Considering Wind Power and Photovoltaic Output Uncertainty [J]. *IEEE Access*, 2021, 9(104): 93-103.
- [21] ZHANG L, TANG C, LIU F, et al. Research on Pricing Mechanism of Electricity Spot Market Based on Multi-Agent Reinforcement Learning (Part II): Decision-making Framework of Pricing Mechanism Combined With Theory and Simulation [J]. *Proceedings of the CSEE*, 2021, 41(3): 1004-1017.

- [22] CHEN Y, WANG Y, KIRSCHEN D S, et al. Model-free renewable scenario generation using generative adversarial networks [J]. *IEEE Transactions on Power Systems*, 2017, 33(99): 3265-3275.
- [23] JIANG C, MAO Y, CHAI Y, et al. Scenario generation for wind power using improved generative adversarial networks [J]. *IEEE Access*, 2018, 6: 62193-62203.
- [24] LIANG J, TANG W. Wind power scenario generation for micro-grid day-ahead scheduling using sequential generative adversarial networks [C] // 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, 2019: 1-6.
- [25] WANG J, ZENG K, ZHANG J, et al. GAN-generated Face Detection Based on Space-Frequency Convolutional Neural Network [J]. *Computer Science*, 2023, 50(6): 216-224.
- [26] FENG H, HUANG C, FENG S, et al. Bidding Model of Generator Considering the Medium and Long-Term contract [C] // Proceedings of the 2019 Annual Meeting of the Power Market Committee of the Chinese Society of Electrical Engineering, 2019: 336-342.
- [27] XU B, ZHU F, ZHONG P, et al. Identifying long-term effects of using hydropower to complement wind power uncertainty through stochastic programming [J]. *Applied Energy*, 2019, 253: 113535.
- [28] WANG D, LIU B, JIA H, et al. Peer-to-Peer Energy Transaction Decision of Prosumers Based on Reinforcement Learning [J]. *Automation of Electric Power Systems*, 2021, 45(3): 139-147.
- [29] KONG D, KONG X, XIAO J, et al. Dynamic pricing of demand response based on elasticity transfer and reinforcement learning [C] // 2019 22nd International Conference on Electrical Machines and Systems (ICEMS), 2019: 1-5.
- [30] HAN D, HUANG W, YAN Z. Deep Reinforcement Learning for Virtual Bidding in Electricity Markets [J]. *Proceedings of the CSEE*, 2022, 42(4): 1443-1455.



HUANG Feihu, born in 1990, Ph.D, is a member of CCF (No. H7742G). His main research interests include reinforcement learning and power big data.



PENG Jian, born in 1970, Ph.D, professor, Ph.D supervisor, is an outstanding member of CCF (No. 22761S). His main research interests include big data and deep learning.