

# 基于 Actor-Critic 强化学习的电力现货 市场单主体卖方决策研究

# 目录

- 第 1 章 绪论.....3
  - 1.1 研究背景和意义.....3
  - 1.2 国内外研究现状.....5
  - 1.3 本文主要内容.....8
- 第 2 章 现货市场与深度学习相关理论知识.....9
  - 2.1 现货市场概述.....9
  - 2.2 电力市场出清机制.....11
  - 2.3 电力市场价格机制.....12
  - 2.4 我国现货市场特点展望.....13
  - 2.5 深化学习模型介绍.....14
- 第 3 章 广东省现货市场调研.....16
  - 3.1 广东电力系统概述.....17
  - 3.2 现货交易情况.....18
- 第 4 章 基于 Actor-Critic 强化学习的电力现货市场单主体卖方决策模型 ...22
  - 4.1 以风电-储能联合系统为代表的新能源卖方决策特性分析 .....22
  - 4.2 决策模型构建.....23
- 第 5 章 现货电力市场仿真.....34
- 第 6 章 结论与展望..... 错误!未定义书签。
  - 6.1 结论..... 错误!未定义书签。
  - 6.2 展望..... 错误!未定义书签。
- 参考文献.....42

# 第 1 章 绪论

## 1.1 研究背景和意义

### 1.1.1 研究背景

电力市场化改革兴起于上世纪八十年代中期，电力市场化改革的基本途径是设法在电力行业中打破垄断，引入竞争，从而使电力商品的价格由市场所决定，其根本目的是为了进一步提高电能的生产和利用效率，从而使之更好地服务于经济社会的发展。电力市场具有开放性、竞争性、计划性和协调性等特点，与普通的商品市场相比，电力市场具有计划性和协调性；与常规的垄断的电力系统相比，电力市场具有开放性和竞争性。我国的电力市场化改革虽然起步较晚，但近十几年来在党中央和国务院的高度重视下，各项改革工作依然取得了显著的成效。2015 年 3 月，中共中央、国务院发布《关于进一步深化电力体制改革的若干意见》（中发[2015]9 号，下称 9 号文），明确指出了我国电力行业市场交易机制缺失、价格关系没有理顺等关键问题，提出了有序推进电价改革、完善市场化交易机制、稳步推进售电侧开放等重点改革任务。新一轮的电力体制改革方案以体制与机制改革并重，力图构建“公平、开放、有序、竞争、完整”的电力市场体系，重点在于还原电力的商品属性，以市场机制优化配置电力资源。2017 年 9 月，国家发展改革委、国家能源局下发《关于开展电力现货市场建设试点工作的通知》特急文件，明确规定：要“组织市场主体开展日前、日内、实时电能量交易，实现调度运行和市场交易的有机连接，促进电力系统安全运行、市场有效运行，形成体现时间和位置特性的电能量商品价格，为市场主体提供反应市场供需和生产成本的价格信号”。这标志着我国现货电力市场建设工作已正式启动。截至 2017 年 12 月，全国在电力交易中心公示的各类售电商已有 2994 家。在各类配、售电商参与市场交易的趋势下，各国有关部门与研究机构已开始逐步深入地对售电侧开放下电力市场的相关决策与运行等一系列问题展开探讨与研究。

### 1.1.2 研究意义

本文立足于中国当前电力市场改革不断推进和人工智能技术飞速发展的现状和背景，对“基于强化学习的发电商中长期电力市场报价决策”开展研究，具有以下三方面的重要科学意义。

#### （1）推动人工智能在电力系统的应用和发展

在中国“互联网+”行动的战略背景下，人工智能将与各行业各领域发生深度融合，持续推动技术进步、效率提升和组织变革。Alpha Go 的成功，说明强化学习方法应用并解决该类型博弈问题的可行性和有效性。电力市场交易和发电商报价决策同样属于博弈问题，在管理、决策等方面具有高度的相似性。强化学习方法能够将机器的精准性和人类的可塑性有效结合，协助提升认知和学习的速度和精度，在电力市场中快速做出最佳决策判断并协同行动，将有利于其在电力系统的进一步应用和发展。

#### （2）保障市场主体在电力市场中的交易利益

电力市场建设的目标是打破电力系统的传统垄断格局，开展有效竞争，构建多元主体积极有序参与的电力市场体系，充分发挥市场在资源配置中的决定性作用。在竞争性电力市场下，各方市场主体面临的关键问题是如何科学合理制定最优的报价策略，获得最大的交易收益。因此，需要报价决策理论和方法指导各方市场主体进行智能报价决策。将强化学习方法应用于电力市场报价决策，实现自主报价及自适应学习，保障发电商在市场竞争中的自身利益，具有重要的理论和实际意义。

（3）促进市场效率提升和电力资源优化配置，支撑电力体制改革目前，中国电力行业处于改革试水的关键时期，各项改革措施尚处于试点阶段。通过对发电商电力市场报价决策进行研究，并基于智能 Agent 对电力市场交易进行迭代仿真验证，能够对电力市场下的各方主体行为进行考察，将有助于识别电力市场机制和规则中的漏洞，提出相应规范性的政策建议，促进电力市场机制和规则的不断完善，提升电力市场的运营效率，实现优化电力资源配置和电力市场的帕累托均衡，有效支撑中国新一轮电力体制改革。

## 1.2 国内外研究现状

### 1.2.1 现货市场

现货电力市场多主体决策研究主要是在现货电力市场环境中,同时考虑所有交易参与者的竞价及出清决策,研究其共同作用下,整个市场竞价过程或均衡状态的建模及仿真分析方法,并力图为市场经济运行模拟、市场规划设计、政策效果估计等相关研究提供分析工具及仿真平台。现货电力市场多主体决策模型通常也被称为现货电力市场竞价模型。国内外有关电力现货市场多主体决策模型构建的研究主要分为两个方面:基于博弈论方法的现货电力市场多主体决策模型以及基于多智能体(Multi-Agent)方法的现货电力市场多主体决策模型。Salem Y A 等(2010)运用供给函数均衡理论(Supply Function Equilibrium, SFE)对单侧(仅存在发电商竞价)日前电力市场中各发电商的竞价策略博弈模型进行构建,并基于该模型对市场的纳什均衡结果展开了讨论。Alberto B 等(2009)运用静态博弈模型(Static Game Model)在考虑系统线路潮流约束的条件下,对单侧日前电力市场的纳什均衡进行分析。在Wang J 等(2011)以及Vijaya K J 等(2014)<sup>[5]</sup>的研究中,每一个发电商都通过判断其他竞争对手竞价策略的概率分布函数来做出本轮最优竞价策略的决策,所有发电商的成本函数及其上一轮竞价时所采用的策略均被视为共同知识。黄仙(2008)基于演化博弈理论方法对单侧日前电力市场中各种群竞价策略的演化过程进行了建模与分析。Ladjici AA 等(2014)构建了基于随机优化理论的电力市场动态博弈模型,模型同时考虑了发电商在远期市场及日前市场中所获得的利润,并通过模型求解,进而对日前市场中的纳什均衡结果进行分析。在含新能源现货市场竞价模型的研究方面,康重庆等(2016)认为新能源发电商(例如风电、光伏等)参与市场对以往以可控常规机组为主的电力市场竞价及其均衡结果(经济运行结果)会产生显著的影响,这主要是由于新能源机组出力的间歇性与不确定性所造成的。综上所述,现货电力市场多主体决策模型的构建问题主要涉及了对博弈论方法以及多智能体方法的运用。由于后者相较于前者而言具有在计算复杂度、决策效果以及对公共知识依赖程度等方面的显著优势,因此已逐渐成为该领域研究的主要发展方向。

## 1.2.2 强化学习

强化学习的基本思想是智能体（Agent）在与环境交互的过程中根据环境反馈得到的奖励不断调整自身的策略以实现最佳决策，主要用来解决决策优化类的问题。其基本要素有策略（Policy）、奖赏函数（RewardFunction）、值函数（Value Function）、环境模型（Environment），学习过程可以描述为如图 1-1 所示的马尔科夫决策过程。

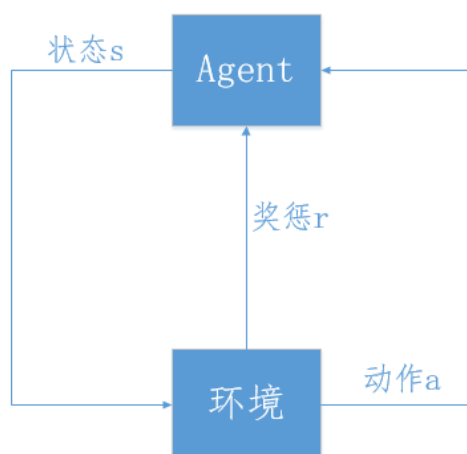


图 1-1 强化学习基本学习模型

学习过程可概括为：智能体（Agent）选择一个动作  $a$  作用于环境，导致状态  $s$  变化为  $s+1$ ，同时产生一个强化信号  $r$ （奖励或惩罚），智能体再次选择下一个动作，动作的选择受强化信号大小的影响。如果  $r$  为正值，以后产生这个行为策略的趋势便会加强。每次选择动作的原则是使受到正的奖励的概率增大。如此反复的与环境进行交互，在一定条件下，Agent 会学习到一个最优的策略。

以深度学习的快速发展和应用为分界线，可将强化学习的研究分为传统强化学习（traditional reinforcement learning, TRL）和深度强化学习（deep reinforcement learning, DRL）两个阶段。

### （1）传统强化学习（traditional reinforcement learning, TRL）

传统强化学习主要涉及基于值函数（value function, VF）的强化学习方法和基于策略梯度（policy gradient, PG）的强化学习方法。

1) 基于值函数的强化学习方法方面：早在 1957 年，Bellman 提出著名的“贝尔曼方程”（Bellman equation）和基于动态规划（dynamic programming, DP）的

强化学习方法，成为后续强化学习方法的重要基础；1988 年，Sutton 系统地构建强化学习的理论框架，同时将动态规划方法和蒙特卡洛方法结合，提出基于时间差分（temporal difference, TD）的强化学习方法；Watkins 和 Rummery 在 TD 强化学习方法的基础上，于 1992 年和 1994 年提出异策略（off-policy）的 Q-learning 算法和同策略（on-policy）的 Sarsa 算法，这两种算法是目前强化学习方法中两种最重要的算法；1994 年，Singh 进一步将 TD 强化学习方法推广，提出基于资格迹（eligibility trace, ET）的强化学习方法，能够有效减少 TD 强化学习方法中的 TD 误差。

2) 基于策略梯度的强化学习方法方面：1992 年，Williams 针对蒙特卡洛随机模拟导致策略梯度方差较大的问题，提出基于 REINFORCE-Baseline 的强化学习方法；2006 年，Kocsis 将置信上限树算法（Upper Confidence bound apply to Tree, UCT）应用于强化学习，为后续 Alpha Go 的诞生和成功奠定重要基础；2014 年，Silver 提出一种基于确定性策略梯度的强化学习方法，相比于随机策略方法，该方法采样数据更少，搜索效率是随机策略方法的十倍以上。

但是，受制于训练数据和计算能力，传统强化学习方法在算法收敛性、可复现性等方面存在一定局限性，因此尚未得到有效和广泛的实际应用。

## （2）深度强化学习（deep reinforcement learning, DRL）

通过结合强化学习的学习机制和深度学习的表达能力，深度强化学习应运而生。2005 年，Ernst 等人提出了拟合 Q 迭代的方法，采用决策树的方式拟合 Q 值，跳出了传统方法使用列表法表征 Q 值的思维定势。同年在 2005 年，Tiedmiller 在 Ernst 拟合 Q 值的思路指导下，采用了多层感知机来做 Q 函数的逼近器，神经网络和强化学习结合了起来。2010 年，Lange 大胆的采用了高维的图片作为系统输入，采用深度自编码网络来对图片进行降维处理，解决了之前强化学习只能解决低维状态空间的限制，使强化学习直接从图像中学习成为可能。2011 年，Abtahi 利用深度神经网络作为 Q 函数的逼近器，并且证明预训练后的深度神经网络有利于深度强化学习的收敛。2015 年，DeepMind 团队采用深度强化学习，用来操控 Atari 游戏，实现了从游戏图像到游戏操作点到点的学习过程，在大量游戏中甚至达到了比人类玩家更好的效果。

深度强化学习进一步发展，现在的深度强化学习已经是多面开花。2015 年，

DeepMind 团队又接连提出了 Double Q Learning 和 Dueling Q Learning<sup>4s</sup> 算法。同年, Lillicrap 提出了 DDPG 算法, 采用深度强化学习方法, 解决连续状态空间和连续动作空间情况的问题。Schulman 提出了 TRPO 算法, 采用深度神经网络, 拟合强化学习的策略函数, 并理论推导了每一次学习置信区间的大小, 使学习能够更快的达到收敛。在 2016 年, DeepMind 团队开发的 Alphago, 通过历史棋谱和自我对弈, 自主学习围棋, 最终战胜围棋冠军李世石。

1.3 本文主要内容

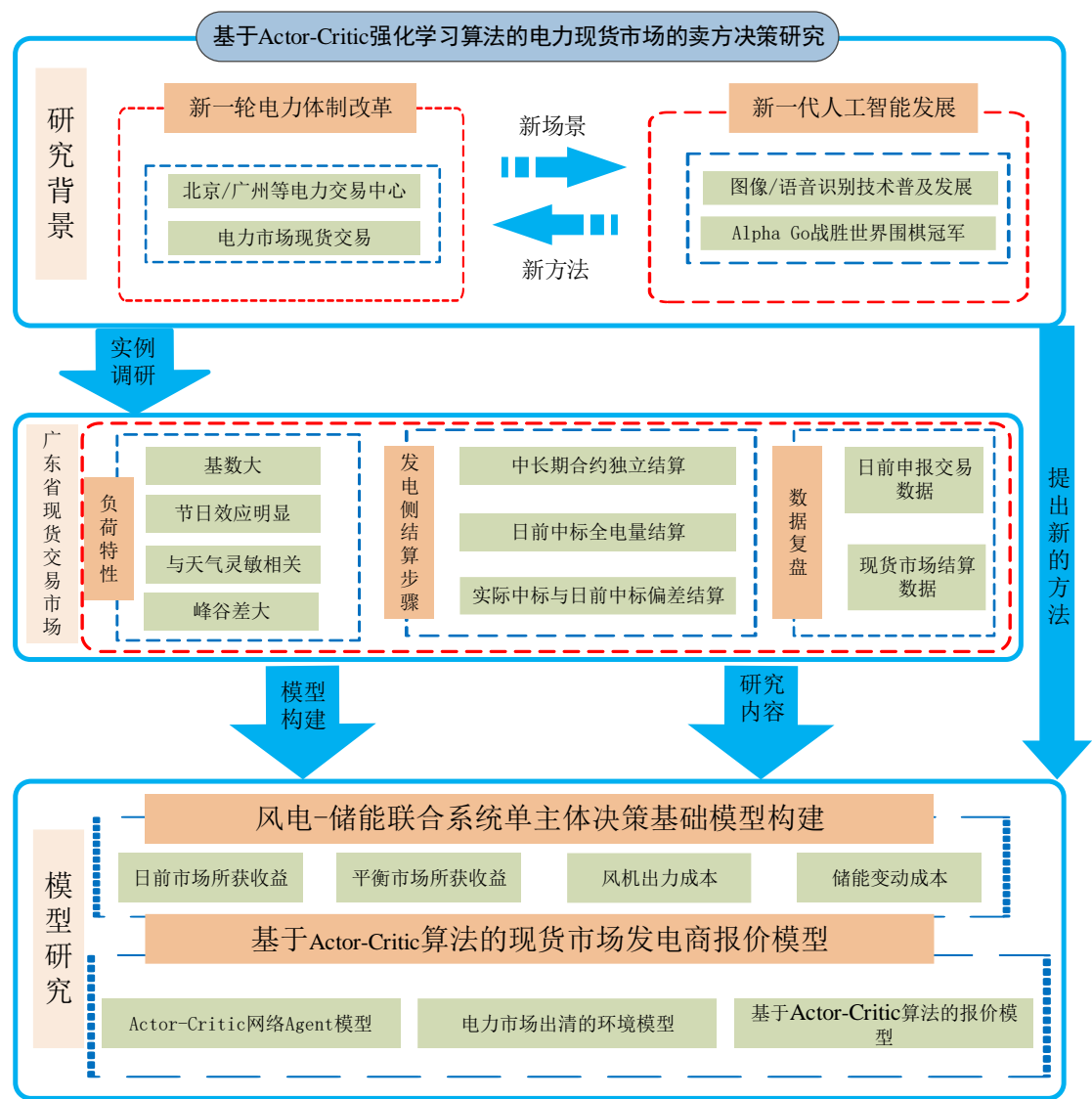


图 1-2 本文研究内容



## 第2章 现货市场与深度学习相关理论知识

### 2.1 现货市场概述

#### 2.1.1 主要作用与基本结构

现货电力市场又称短期电力市场，其通常是针对实际用电时刻提前数（十）分钟，数小时，甚至一天所组织与开展的电能量实物交易市场。根据相关文献中的相关分析，现货电力市场的主要作用与意义在于：

（1）可在一个适合的时间前提量上形成与电力系统物理运行相适应的、体现市场成员意愿的优化交易计划；

（2）能够促进电能量交易的充分竞争，实现电力资源的高效、优化配置；

（3）能够充分发挥市场价格形成的功能，真实反应电力商品短期供需关系和时空价值，为有效的投资和发展提供真实的价格信号；

（4）为市场成员提供一个修正其中长期发、购电计划的交易平台，减少系统的安全风险；

（5）为电力系统的阻塞管理提供调节手段和经济信号，真实反应系统的阻塞成本，保证电网的安全运行等。

为实现上述目标，通常情况下，完善的现货市场体系应包含两个重要的市场阶段，即日前市场（Day-Ahead Market）与平衡市场（Balancing Market）。介于日前与平衡市场的现货市场常被称为日内市场，由于日内市场的交易规模在整个现货市场体系中所占比重很小，因此，常常在相关的现货市场研究中被忽略。日前市场与平衡市场的同时存在是非常必要与合理的，一方面，对于那些反应速度较慢，需要提前做好发电计划的机组而言，日前市场的存在为高效可靠地对其进行出力调度提供了必要条件；另一方面，通常情况下，对负荷或各类新能源机组出力的短期预测是存在误差的，这使得电力系统的有功功率在实时往往会存在不平衡的风险，平衡市场引入了促使系统中反应迅速的各类资源在实时改变其日前调度结果的竞争机制，进而可以高效可靠地平衡电力系统的实时有功功率，从而为保障系统安全可靠与经济运行，保障电能质量起到了重要的促进作用。

## 2.1.2 组织方式与决策主体

现货电力市场是通过集中竞价的方式组织交易的,因此,在现货电力市场中,除了交易参与者外,还应包括交易的组织者,即负责组织集中交易及运行市场出清的机构或部门。在不同国家或地区的现货电力市场中,扮演交易组织者角色的机构常存在差异美国 PJM 电力市场及北欧电力市场均在其现货市场中均允许由各类发电商,配、售电商及大用户进行双侧竞价,因此,这两类市场的开放性程度较高,是常被各类研究所借鉴与引用的主流市场模式。另外,为了便于本文后续论述工作的简便与清晰,我们不加区分地认为在整个现货电力市场体系中,担负交易组织者职责的机构均为 ISO,其所运行的日前及平衡市场出清机制均是基于阻塞管理方式所进行的。日前市场的开展应当比电力的实际交付提前一天(通常是在前一天的中午)进行。在日前市场中,发电商针对次日各交易时段(例如,24 或 48 个交易日)一次性向 ISO 申报相应数量的出力供给曲线,该曲线又称发电商的报价或竞价曲线,通常是由分段的有功出力与相应的最低可接受价格所组成的。若该日前市场是双侧竞价日前市场,则在发电商向 ISO 报价的同时,市场中各配、售电商及大用户也针对次日各交易时段一次性向 ISO 申报相应数量的负荷需求曲线,该曲线又称售电商或大用户的报价或竞价曲线,是由分段的有功负荷与相应的最高可接受价格所组成的。ISO 在收到供求双侧的申报计划后,运用一定的市场出清方法,例如前文所提及的基于阻塞管理的前日市场出清方式,便可得到各发电商,配、售电商及大用户次日各时段的出力、负荷出清结果(日前经济调度计划)以及相应的前日出清结算价格。平衡市场应当比电力实际交付时刻提前数分钟至数十分钟开展,并且是最后一个平衡电力生产与消费的现货市场阶段<sup>[34]</sup>。电力系统中越是大量存在实时出力具有间歇性、不确定性及低可调度性特点的资源(例如,以风电、光伏为代表的各类新能源发电商),平衡市场的作用便越是显著。平衡市场中交易参与者的申报方式及组织者的市场出清机制与日前市场类似,其出清结果包括系统的实时经济调度结果及相应的平衡出清结算价格。然而,不同国家及地区的平衡电力市场在结算方式方面存在差异。日前及平衡市场的申报及出清方式将在后文中详细阐述。综上所述,在现货电力市场中,交易参与及组织者均是重要的经济行为决策主体一方面,各类发电商,配、售电

商等交易参与者应对其自身交易申报计划（通常为报价曲线）进行决策，以期实现在现货市场中交易获利的目的；另一方面，ISO 等交易组织者应在各交易参与者集中申报的前提下，负责对现货市场出清进行决策，从而得到相应的经济调度结果及市场出清电价。显然，由于现货市场是由上述主体所共同组成的经济系统，因此，整个市场以所有主体共同合理决策的作用下实现优化配置电力资源为最终目的。

## 2.2 电力市场出清机制

现货交易出清机制与现行的中长期交易出清机制有很大不同。目前中国中长期交易电量不分解到日，具体执行时由调度统筹安排。因此中长期交易出清结果对于系统运行的影响较小，出清模型也比较简单，主要考虑的是报价因素。现货交易的出清结果与调度运行直接相关，日前市场出清直接生成预调度计划，实时市场出清决定实时调度安排，因此出清模型需同时考虑报价因素和系统物理模型。

在只考虑报价因素的情况下，交易机构根据市场主体提供的报价曲线进行叠加得到总发电和总用电价格曲线，这两条总发用电价格曲线的焦点即为市场出清点，出清点左侧的交易全部成交。为了协调交易出清与电网调度运行的关系，现货市场的出清模型除了市场主体报价外，还需考虑系统平衡、机组能力及网络约束。在日前市场，选择分散式市场的地区一般具有系统调峰资源好、阻塞较轻、电力资源供应充足等特点，可以选择不带物理模型，按照报价情况出清，由调度对成交的交易进行安全校核，或者不校核直接交由实时市场调节。选择集中式市场的地区一般具有系统调峰资源薄弱、阻塞较严重、电力资源供应紧张等特点，出清时应考虑最为精密的物理模型。具体做法是通过安全约束机组组合（SCUC）和安全约束经济调度（SCED），确保出清结果能够满足系统安全稳定运行需要。为充分反映各类设备的真实运行状态，发电商需上报其实际物理参数，如机组运行的爬坡速率、最大最小出力、连续开停机时间约束等技术参数。在实时市场，出清结果直接用于系统的平衡调度与阻塞管理。因此，无论是分散式市场还是集中式市场，都以最精确的系统物理模型为约束，按价格最低为原则出清。实时出清的具体运作机制与考虑物理模型约束的日前出清大体相同，但是由于机组组合在日前已经确定，实时市场只需运行 SCED，不需要 SCUC。

## 2.3 电力市场价格机制

电力现货市场交易周期较短，网络约束对交易结果也有一定影响，因此价格机制相对复杂，一是要反映电价在不同时段随需求波动的特点，实现市场发现价格的目标；二是要反映在网络条件影响下电价随空间变化的情况，依靠价格引导减轻网络阻塞；三是要维护市场稳定，防止市场主体的投机行为。

电力现货市场价格机制主要有统一边际电价、节点边际电价、分区边际电价 3 种。现货交易中，市场主体对不同时段进行报价，不同时段出清价格即可反映电价随时间和需求的变化。由于这 3 种价格机制都是采取边际电价成交而非报价成交的方式，发电企业会趋于按照各自边际成本报价以保证交易成功，因此都可以在一定程度上防止发电企业哄抬电价的投机行为。

这 3 种价格机制的适用范围有所不同，各有其优缺点。统一边际电价机制主要用于没有网络阻塞的系统，系统内所有成交交易都按照出清点对应的价格成交，电价不随空间变化。这种方法虽然操作简单，但是当系统内出现阻塞时，阻塞地区的高价机组可能会被调用，造成统一边际电价上升，抬高系统总购电成本。节点边际电价主要用于网络阻塞比较严重的系统。节点边际电价可以反映不同位置电力供需形势和网络阻塞情况，为电厂、电网的规划布局提供参考，局部阻塞也不会拉高全系统的购电成本。但是节点边际电价算法较为复杂，必须根据实际网络模型精确计算。分区边际电价主要适用于网络阻塞发生比较确定的系统。按照网络阻塞将系统分为若干分区，采用分区边际电价对各分区进行结算，可以在一定程度上反映电价随空间变化的情况，某一分区的边际电价不会影响其他分区的购电成本，其操作难度介于前两者之间。在现货市场的建设过程中应因地制宜选择价格机制。存在网络阻塞的地区应根据具体的阻塞情况，尽早建立分区边际电价或节点边际电价机制，在 SUCU、SCED 程序中集成节点边际电价计算功能。对于目前不存在阻塞的地区，可以采取从简到繁的方法，先采用统一出清电价机制，后期有需要时再过渡到分区边际电价和节点边际电价，也可以采用一步到位的方式，先建立节点边际电价机制，然后根据网络特性将其简化为分区边际电价或节点边际电价。由于节点边际电价机制与现行电价机制差异非常大，其复杂的算法也需要有成熟的技术支持系统支撑，因此具体实施时应该采用试点先行的方式。

## 2.4 我国现货市场特点展望

本节在结合《关于开展电力现货市场建设试点工作的通知》文件中的各项基本要求以及借鉴前文所述各项国外成熟经验的基础上,对我国未来所逐步形成的现货电力市场体系主要特点进行展望如下(本文不涉及对辅助服务的探讨):

(1) 针对一个交易日而言,应陆续展开日前、日内及平衡(实时)市场电能交易,其中,日前及平衡市场是必不可少的两个现货市场阶段。多阶段开展现货市场的目的在于确保电力系统有功功率的实时平衡,即以市场方式完成系统三次调频的主要任务。

(2) 现货市场中应存在多元化的交易参与者主体,并且涵盖买、卖双侧竞争。常规发电商,各类新能源发电商,配、售电商以及各类大用户等均应成为主要的市场交易成员。优先接纳新能源发电、鼓励功率可灵活调节资源多样化参与市场等各项相关政策及措施应在现货市场逐渐发展的过程中不断制定、落实与完善。

(3) 应建立基于阻塞管理的现货市场出清机制,从而可在网络安全约束前提下,形成分时、分节点或区域的电能商品价格,例如 LMP 等。

(4) 现货市场首先应通过分散试点的方式,以省或现有电力调度控制区为单元分别独立开展。在初建阶段,各省(区)现货电力市场应在原有调度范围的基础上独立运行,以实现通过市场方式在省(区)内优化配置电力资源的经济目标。待各省(区)现货电力市场发展相对完善与成熟后,应鼓励市场逐步联合,从而形成若干基于特高压输电通道的跨省(区)现货电力市场,以实现通过市场方式在省(区)间,乃至全国范围内优化配置电力资源的经济目标。另外,在跨省(区)现货电力市场发展相对完善与成熟后,可鼓励有条件的市场通过跨国输电通道与“一带一路”上相邻国家进行联合,从而形成若干跨国现货电力市场,以实现通过市场方式在更大范围内优化配置电力资源的经济目标。

综上所述,本文认为,我国未来所逐步形成的现货电力市场体系应具有多阶段,多元主体,省(区)内独立运行,省(区)间协调联络等时空兼备特点。现货市场体系的有效运行,一方面,能够确保我国电力系统有功功率的实时平衡,完成系统三次调频的主要任务;另一方面,能够最终实现省(区)内、省(区)间、全国甚至更大范围内的电力资源优化配置目标。

## 2.5 深化学习模型介绍

### (1) 强化学习基本模型

强化学习的目的就是求解马尔可夫决策过程的最优策略。类似于构成 MDP 的元素，强化学习过程也由几个要素组成：决策、回报函数、值函数和环境模型。下面详细介绍它们的具体定义：

a. 策略 (Policy)，也称为决策函数，定义了 Agent 在特定时间特点环境下的行为方式  $S \rightarrow A$ ，是一个从状态到动作的映射。

策略可分为两类：确定性的策略 (Deterministic Policy)，随机性的策略 (Stochastic Policy)。确定策略就是某一状态下的确定动作  $a = \pi(s), a \in A, s \in S$ ，随机策略用概率描述，即某一状态下执行某动作的概率  $\pi(a|s) = P(a_t = a | s_t = s)$ 。决策是强化学习的核心部分。

b. 回报函数 (Reward Function)  $r$ ：是一个从环境状态到实数的映射  $S \rightarrow R$ ，可视为标量值，通常用字母  $r$  表示。

回报函数可以理解为环境执行 Agent 发出的动作后，返回给 Agent 的一个信号，用来评价所选动作的好坏。Agent 可以根据回报函数的值来调整选择动作的策略。若回报信号为一个正数，则表明之前动作是好的，之后出现这个动作的概率也会提高，反之，则出现此动作的概率将会降低。

c. 值函数 (Value Function)：如果策略只考虑瞬时回报，很可能会陷入局部最小，故在选择行为策略过程中，考虑到回报目标的长远性和强化学习的延迟回报问题，在策略和瞬间回报之间构造值函数 (又称状态的效用函数)，用  $V$  表示。如  $V(s_t)$  表示状态  $s_t$  时的值函数，指在状态  $s_t$  时采用动作  $a_t$  后得到的所有未来回报的累积，即所有将来奖赏值通过折扣因子  $\gamma[0,1]$  作用后的总和：

$$V(s_t) = E \left( \sum_{i=0}^{\infty} \gamma^i r_i \right) \quad (1)$$

其中， $r_t = r(s_t)$ ，为  $t$  时刻的瞬时回报。

d. 环境模型 (Environment)：强化学习将试探学习看成获得经验的一个过程。环境会接收 Agent 执行的一系列的动作，并对动作的好坏进行评价，转换成一种标量信号即 Reward 反馈给 Agent。同时，环境还向 Agent 提供它所处的状态信

息。

强化学习作为一种特殊的机器学习方法，具有如下特点：

- 1) 强化学习是通过与环境不断的试错交互来进行学习，没有教师信号，不要求先验知识。强化信号的反馈有延时，不能立即返回。
- 2) 强化学习可以在线使用。
- 3) 强化学习可以应用于不确定性环境。
- 4) 强化学习的体系结构可扩展。目前，强化学习系统已扩展至规划合并、智能监督学习和结构控制等领域。

## (2) Actor-Critic 模型

Actor-Critic 结构也称为行动者-评论家学习算法，是一种重要的强化学习算法，其结构如图 2-1 所示。Actor-Critic 方法融合了以值函数为基础的 Critic-only 和以动作选择为基础的 Actor-only 两种方法的特点，同时对马氏决策过程的值函数和策略函数进行逼近，因此具有更好的控制效果。

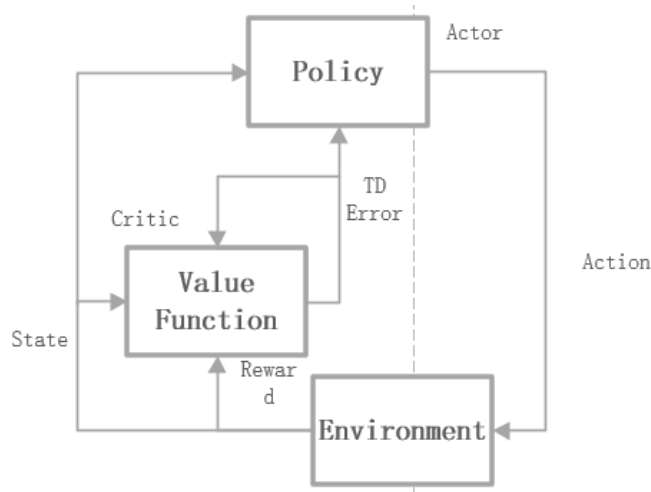


图 2-1 Actor-Critic 算法结构示意图

Actor-Critic 学习模型最主要分为两大部分：行动者（Actor）部分和评价器（Critic）部分。Actor-Critic 学习可对值函数和策略函数进行逼近，其中策略估计由 Actor 部分通过策略梯度估计方法进行梯度下降学习来实现；而值函数估计由 Critic 部分采用 TD 学习算法实现。图 2-1 给出了 Actor-Critic 学习模型的体系结构：对于状态  $s$ ，行动者根据当前策略选择动作  $a$ ，状态  $s$  接受动作  $a$  的作用后，转移到状态  $s+1$ ，同时产生一个回报信号  $r$ ；状态  $s$  和回报信号  $r$  作为评论家的输入，其输出为值函数的估计，并产生一个 TD 误差信号，用于评论家和行动

者网络的更新学习，对选择的动作做出评价，以修正评论家的动作选择策略。Critic 对值函数估计采用如下学习算法：

$$w_{t+1} = w_t + \alpha[r_t + \gamma V(s_{t+1}) - V(s_t)]z_t \quad (2)$$

式中， $w_t$ 为权值向量， $s_{t+1}$ ， $s_t$ 分别为  $t+1$  和  $t$  时刻的状态， $V(s_t)$  为  $t$  时刻的值函数估计， $z_t$ 为适合度轨迹，定义为：

$$z_t = \gamma \lambda z_t + \frac{\partial V(s_t)}{\partial w_t} \quad (3)$$

设 Actor 输出为

$$\bar{y}_t = f(s_t, w_t) \quad (4)$$

经高斯分布函数后，实际控制量输出为

$$p(y_t) = \exp[-\frac{(y_t - \bar{y}_t)^2}{2\sigma_t^2}] \quad (5)$$

其中，

$$\sigma_t = \frac{k_1}{1 + \exp[k_2 V(s_t)]} \quad (6)$$

其中， $k_1 > 0, k_2 > 0$ 为系数。

Actor 学习算法如下：

$$\frac{\partial J_\pi}{\partial w} = \frac{\partial J_\pi}{\partial \bar{y}_t} \frac{\partial \bar{y}_t}{\partial w} \approx \hat{r}_t \frac{y_t - \bar{y}_t}{\sigma_t} \frac{\partial \bar{y}_t}{\partial w} \quad (7)$$

其中， $J_\pi$ 为学习性能指标， $\hat{r}_t$ 为内部回报，其计算式为：

$$\hat{r}_t = r(t) + \gamma V(t+1) - V(t) \quad (8)$$

Actor 的权值学习算法作为一种近似的策略梯度估计方法，以极大化回报性能指标为学习目标。值函数估计的正确性能够直接影响 Actor 的行为选择和参数更新等，是算法求解学习控制问题的关键一环。

### 第 3 章 广东省现货市场调研

随着《广东省现货电能量市场交易实施细则》等一系列相关文件的出台，2019 年 5 月 20 日，南方（以广东起步）电力现货市场发布国内首次电力现货交易结算结果，包括 190 台发电厂机组、123 家售电公司、3 家大用户在内的广东全部市场主体参与本次结算，期间现货市场和电力系统的运行保持稳定、有序。15 日



用户侧统一结算价格最低 82 元/MWh，最高 362 元/MWh，平均 261 元/MWh。发电侧平均价格最低 70 元/MWh，最高 687 元/MWh，平均 263 元/MWh，总成交量 8.61 亿千瓦时。本文将以广东省为调研对象，对市场数据进行深入分析。

### 3.1 广东电力系统概述

近年来，随着广东省电力体制改革的深入推进以及中长期及现货市场的改革与试点工作的开展，其电力负荷的增长和发展趋势呈现出明显的负荷特性，现总结如下表所示。

表 3-1 负荷特性

基 数 大	广东省经济发达，电力需求旺盛，负荷和电量均稳居全国省级电网首位。 2017 年最高统调负荷 10858 万千瓦，同比增长 8.5%。
节 日 效应明显	广东注重历史文化和传统民俗，各类社会活动对电力负荷有明显影响。
与 天 气 灵 敏 相 关	气候以亚热带季风气候为主，降水在年内分配不均，台风频繁，气象条件的变化对负荷需求也有十分显著的影响。空调负荷超 2500 万千瓦，用电负荷存在典型的气温驱动型特性，在持续高温或寒潮天气、台风天气的影响下，负荷波动可达千万级。
峰 谷 差大	日峰谷差一般占最高负荷的 30%-40%。2018 年最大峰谷差达 4607 万千瓦（不考虑节假日和台风等特殊时期），为系统调峰带来了挑战。峰谷交替期负荷变化速率快，达到 300 万千瓦/分钟，调频难度大。

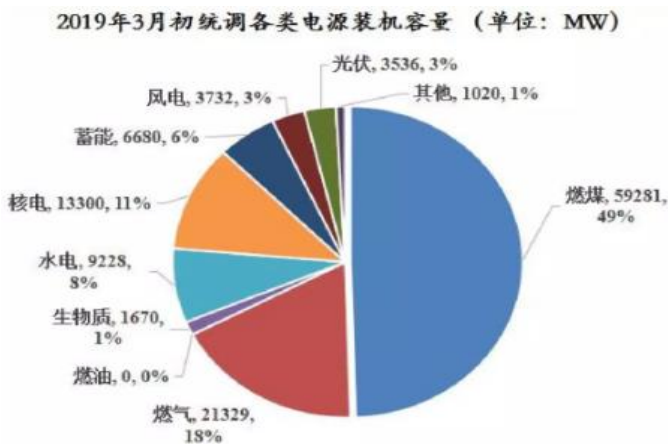


图 3-1 电源结构及特点



图 3-2 不同电源类型及特点

## 3.2 现货交易情况

广东省电力市场发电侧结算步骤见下表。

表 3-2 电量结算方式

①中长期合约	中长期合约独立结算
②日前报量报价	日前中标全电量结算
③实时偏差	实际中标与日前中标偏差结算

本文结合调研收集的相关数据，对广东省现货交易进行复盘。

### A.总盘数据：

中长期电量包含双边协商，年度挂牌及月度竞价电量，总电量 5.82 亿千瓦时；

曲线分解按照市场用户总负荷系数比例统一分解，日前申报电量 5.62 亿千瓦时；



图 3-3 现货市场日前申报交易数据

### B.分解数据



图 3-4 现货市场结算数据

结果分析：

- 中长期交易电量占 103%，且分解曲线是按照统调负荷曲线作为参考的，现货价格波动对价格影响不大；
- 用户侧综合分时电价，峰谷价差很小（中长期价格锁定价格）；
- 用户侧日前、实时分时电价，峰谷价差也不大（0.28 元）。

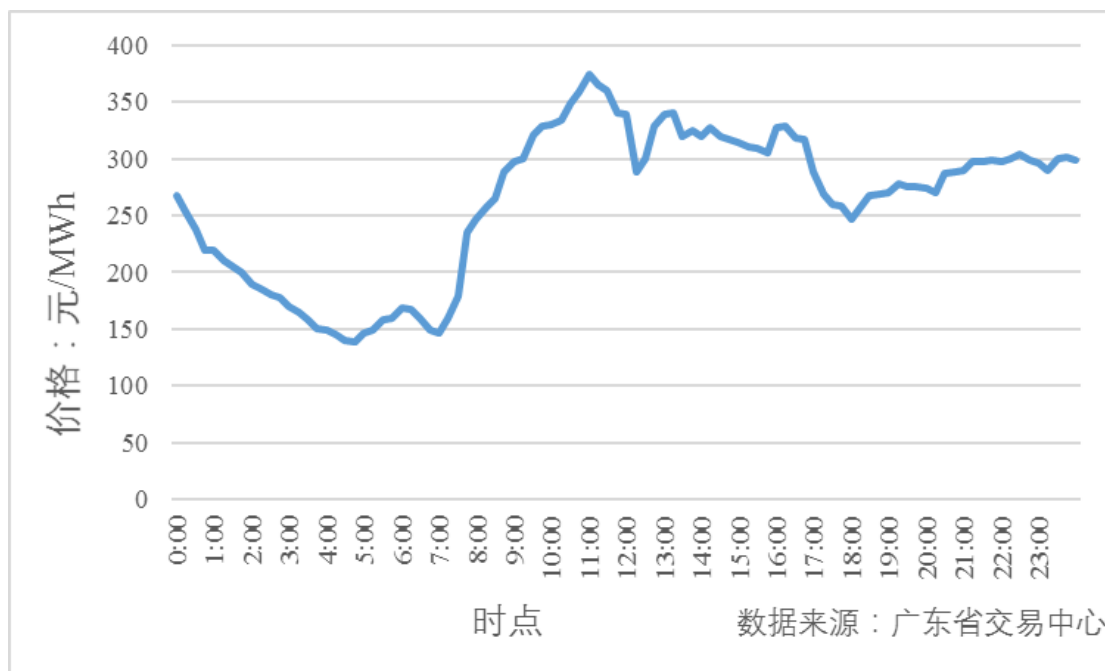


图 3-5 用户侧统一结算价格曲线

16 日用户侧统一结算价格最低 139 元/MWh，最高 374 元/MWh，平均 295 元/MWh。发电侧平均价格最低 70 元/MWh，最高 712 元/MWh，平均 297 元/MWh，总成交量 9.42 亿千瓦时。

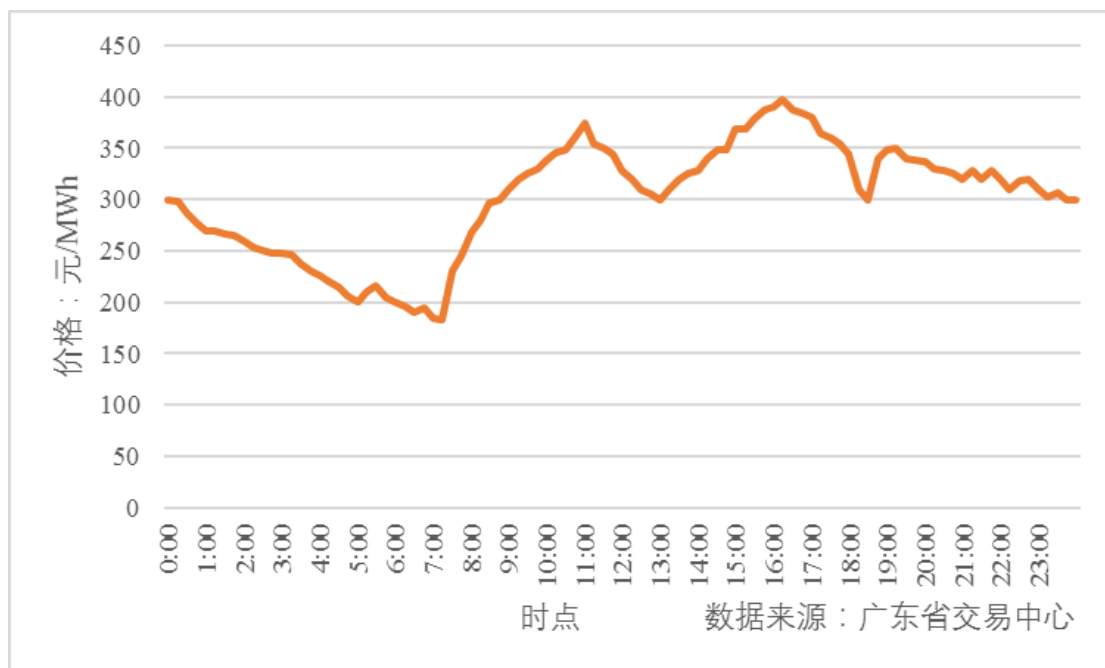


图 3-6 发电侧平均价格曲线

15,16 日的试结算结果与 17 年最开始的中长期市场月度竞价结果是类似的，且相对之前模拟的出清结果明显更高，一定程度上能反映市场供求关系。可以看

出市场主体，尤其是发电侧在试结算日的报价是有准备的，普遍为了求稳报出偏低的价格以求中标。

从这轮试结算中可以看出，由于现货市场中是根据现货市场成交量价执行物理交割，而中长期交易量价是通过差价合约进行金融结算，用于对冲现货市场风险。因此我们可以理解发电厂们在此次日前申报中“赴死”报低价，是为了保现货中标电量，发电厂敢在现货报赔本价格成交，是因为有中长期成交量价作为后盾，保价差收益。

小结：首次电力现货交易结算平均节点电价 0.263 元/度。用户平均节点电价 0.263 元/度，夜间低谷时期平均节点电价 0.082 元/度电，负荷高峰时期平均节点电价 0.362 元/度电。”此次结算选取 5 月 15 日、16 日两天为结算运行日，14 日、15 日两天为交易申报日，20 日为结算结果发布日。由发电商自主申报拟出售的电量和价格，由购电用户自主申报拟购买电量，经过交易平台竞价撮合，并通过电网运行安全检验，每 15 分钟形成一个电价，不同地方、不同时段用电价格不同。发电商以什么价格发多少电、用户以什么价格购电，由市场这只“无形的手”，按照市场规律决定。

长期参与我国电力市场化改革的知名专家、清华大学夏清教授谈到，从国际成熟电力市场看，电力现货市场是电力价格发现和资源优化配置最重要的环节，现货市场功能是为了让分时、分位置的价格信号更充分地反映电力供需情况，实现电力资源时空优化配置。现货市场建设的目的不能简单认为是降电价，而是让价格反映电力在时间与空间上的供求关系。现货市场反映的是电力电量市场供求关系，改变了以往中长期交易只考虑电量平衡，重构了中长期电力交易体系，将极大地降低社会为电力平衡付出的成本。通过培养用户科学友好的用能习惯，获得降低社会能耗水平、提升资产利用效率、新能源消纳能力的长远改革红利。

## 第4章 基于 Actor-Critic 强化学习的电力现货市场单主体 卖方决策模型

### 4.1 以风电-储能联合系统为代表的新能源卖方决策特性分析

由于对常规发电商或风电厂等未联合功率可灵活调节资源的新能源发电商现货电力市场单主体决策问题的研究在近年来较为常见，因此，在本章中，“价格接受者”型卖方现货电力市场单主体决策模型的构建是围绕一个风电-储能联合系统（WindFarm-Energy Storage System, WF-ESS）所展开的。风电（以及其他常见的新能源发电）在未联合各类功率可灵活调节的资源时，其实时自然出力具有显著的间歇性、不确定性及低可调度性等特点，这就使得其往往成为造成电力系统有功功率实时偏差的主要“根源”之一。由于储能设备具有快速充放电的功能，这使得若将风电厂与储能设备联合成同一个主体，则这个联合主体（即联合系统）便可在各平衡市场阶段自发地对自身（联合）实时出力进行一定程度的调整，从而一方面，为平抑上述功率实时偏差，消纳风电出力起到促进作用；另一方面，也使得联合系统具备了在平衡市场阶段的策略性交易计划申报能力。“价格接受者”型风电-储能联合系统在各级现货市场中往往只针对各交易时段申报其联合出力计划（或出力调整计划），这些出力计划或是因为政策原因（例如，新能源优先上网政策），或是因为零报价（或极低报价）原因，会被市场优先全额接纳。本章在建模过程中拟采纳上述关于风电-储能联合系统交易计划申报方式的观点，因此，本章建模问题的关键在于如何对“价格接受者”型风电-储能联合系统在各级现货电力市场中的联合出力计划进行优化决策。另外，本章认为现货市场体系只是由日前及平衡市场两个阶段组成，时前等日内市场阶段均被忽略，且本章中也同时忽略了储能设备提供辅助服务的可能。

另外，为了问题分析的简便且不失一般性，除前文基本说明外，还应对如下不确定因素及各项假设予以进一步明确：

（1）关于各项不确定性因素：一方面，在现代电力市场体系中，市场出清是在全部参与者交易计划申报之后所进行的，这意味着参与者在决策阶段并未掌

握本轮市场出清结果的确切信息；另一方面，由于风电实时自然出力具有显著的间歇性、不确定性及低可调度性等特点，这使得决策阶段距实际发、用电阶段越远，风电实时自然出力的确切信息就越难把握。因此，对于“价格接受者”型风电-储能联合系统等新能源发电商而言，其参与现货电力市场所面临的主要不确定性因素可以概括为：日前、平衡市场的出清价格以及风电实时自然出力。

（2）假设对于本章所研究的风电-储能联合系统而言，其内部仅包含一个风电机组和一个广义概念上的储能装置。

（3）对于一个具体的交易日而言，一共存在一个日前及  $T$  个平衡市场。其中，日前市场于前一天中午开展，同时涉及该交易日  $T$  个时段的申报及出清问题； $T$  个平衡市场分别针对每个时段提前十几分钟开展，各自解决对应时段的功率实时平衡问题。另外，平衡市场中采用单一价制结算方式。

（4）风电实时自然出力符合时间序列特点，即距预测时点越近，预测精度越高；另一方面，平衡市场的开展已非常接近电能的实际交付时间。因此，在第  $t$  ( $1 \leq t \leq T$ ) 个平衡市场决策及出清阶段，假设针对  $t$  时段的风电实时自然出力已变为确定信息，即  $t$  时段开始时，对其相应的风电实时自然出力的预测精度已经达到较高水平。

## 4.2 决策模型构建

在现货市场环境中，单个交易参与者的主要经济目标是通过采用科学的决策方法，生成合理的交易计划申报策略，从而追求自身利润最大限度的提高。当现货市场交易规模较大，或单个交易参与者的供给或需求容量相对于市场规模而言非常小时，该参与者的市场力往往较小，可被近似认为是市场的“价格接受者”，即其交易申报行为对市场的出清价格及总交易规模没有影响（或影响微小，可忽略不计）。此时，以单个“价格接受者”型参与者为对象研究现货市场单主体（交易计划申报）决策建模问题更具有实际意义。这是因为，一方面，在较大规模的市场中研究所有参与者共同竞价的多主体决策问题在技术上是不可行的；另一方面，市场出清电价等在决策阶段应视为具有不确定性特点的环境因素往往是独立于单个“价格接受者”型参与者市场行为的，这大大简化了单主体决策模型的构建及仿真难度。因此，本文将从“价格接受者”型卖方单主体角度研究其现货电

力市场经济行为决策模型的构建及仿真问题。

同时根据现货电力市场和发电商报价的场景，发电商申报电量和申报电价均为连续型变量，基于值函数的强化学习方法不适用于现货电力市场发电商报价模型。同时，为实现现货电力市场发电商报价的增量迭代计算和“在线学习”，将基于 Actor-Critic 的强化学习方法用于电力现货市场发电商报价模型中。

#### 4.2.1 风电-储能联合系统现货电力市场决策过程

结合各项相关说明及假设，本章“价格接受者”型风电-储能联合系统参与现货市场交易的单主体决策过程可概要概括为：对于一个交易日而言，“价格接受者”型风电-储能联合系统先后参与了一个日前市场与  $T$  个平衡市场的交易。在日前市场决策阶段，日前及所有平衡市场出清电价、次日各时段风电实时自然出力均是不确定的。联合系统可通过一定的决策方法申报次日各时段的联合出力计划，并在日前市场出清后获得等于日前出清价格与联合出力申报量乘积的日前市场收益。随着时间推移，针对该交易日的  $T$  个平衡市场陆续开展，在第  $t$  ( $1 \leq t \leq T$ ) 个平衡市场决策阶段，一方面， $t$  至  $T$  时段平衡市场出清电价、 $t+1$  至  $T$  时段风电实时自然出力是不确定的；另一方面，根据假设（4），此时针对  $t$  时段的风电实时自然出力已变为确定信息，且往往与其日前预测量之间存在较大差异。联合系统根据更新的风电实时自然出力信息不仅可以最终确定  $t$  时段风电机组的实时出力，同时也可以灵活确定储能设备的实时充放电状态及功率。这样一来，在联合系统自身调度得当的前提下，其即可通过内部调节以平抑自身风电出力偏差，降低系统实时联合出力的不确定性，又可向平衡市场主动提供联合出力上、下调节服务，即主动参与平衡市场交易计划申报。联合系统在第  $t$  个平衡市场出清后，可获得或支付等于平衡出清价格与自身联合出力上、下调节量乘积的收益或费用。

#### 4.2.2 风电-储能联合系统单主体决策基础模型构建

如前文所述，风电-储能联合系统在日前市场决策阶段可获得次日各时段的联合出力计划，其在各平衡市场决策过程中除了最终确定各对应时段的实时风电



出力之外，还可以灵活确定对应时段储能设备的实时充放电功率，这样一来，在联合系统自身调度得当的前提下，不仅可以通过内部调节平抑自身风电出力偏差，即降低系统实时联合出力的不确定性，还可以向平衡市场主动提供联合出力上、下调节服务，即主动参与平衡市场交易计划申报。对于一个完整的交易日而言，“价格接受者”型风电-储能联合系统在参与完所有日前及平衡市场后，所获的总利润为：

$$R_{WF-ESS} = \sum_{t=1}^T \lambda_t^* P_t^{da} + \sum_{t=1}^T \lambda_t^B P_t^B - C_W \sum_{t=1}^T W_t - C_E \sum_{t=1}^T E_t$$

式(3-1)中， $\sum_{t=1}^T \lambda_t^* P_t^{da}$ 项表示联合系统在日前市场中因申报联合出力计划 $P_t^{da}$  ( $1 \leq t \leq T$ )而在日前出清电价 $\lambda_t^*$  ( $1 \leq t \leq T$ )下所获得的收益； $\sum_{t=1}^T \lambda_t^B P_t^B$ 项表示联合系统在各平衡市场中因申报联合出力调节计划 $P_t^B$  ( $1 \leq t \leq T$ )而在平衡出清电价 $\lambda_t^B$  ( $1 \leq t \leq T$ )下所获得或支付的收益或费用，其中，若 $P_t^B \leq 0$ ，则对应于联合系统在第 $t$ 个平衡市场计划提供联合出力的向下调节，若 $P_t^B \geq 0$ ，则对应于联合系统在第 $t$ 个平衡市场计划提供联合出力的向上调节； $C_W \sum_{t=1}^T W_t$ 项表示联合系统中风机为实现实时出力 $W_t$  ( $1 \leq t \leq T$ )，而在成本系数 $C_W$ 下所产生的发电(变动)成本， $C_W$ 通常为常量； $C_E \sum_{t=1}^T E_t$ 项表示联合系统中储能设备因实时充、放电功率 $P_t^C$ 及 $P_t^D$  ( $1 \leq t \leq T$ )导致 $t$ 时段末残余能量为 $E_t$  ( $1 \leq t \leq T$ )时，在成本系数 $C_E$ 下所产生的储能(变动)成本， $C_E$ 通常也为常量。显然，在日前市场阶段，联合系统的主要决策变量为 $P_t^{da}$  ( $1 \leq t \leq T$ )；而在第( $1 \leq t \leq T$ )个平衡市场决策阶段， $P_t^B$ 、 $W_t$ 、 $P_t^C$ 及 $P_t^D$ 均需要联合系统进行决策。

虽然(针对同一个交易日的)不同市场决策阶段的主要决策变量有所不同，但这并不意味着各决策阶段之间不存在联系。事实上，若不考虑后续阶段，则当前阶段的决策结果很可能导致后续阶段决策效果欠佳，甚至无解，而通常情况下，风电-储能联合系统(或其他参与者)所追求的目标应为所有市场阶段所获利润的总和。因此，本文认为，追求总利润最大化(基本目标)的风电-储能联合系统不应将各市场阶段相互割裂并独立对待，而应具有前瞻性视野(Prospective Point of View, PPV)，即在当前的决策阶段，应同时考虑对本阶段变量的最终决策结果以及对随后若干阶段变量的初步决策结果进行求解。后文中，经最终决策所得到的变量结果通常会加上一个上角标“\*”以表示最终值。

### 4.2.3 Agent：Actor 策略网络和 Critic 价值网络模型

由于对常规发电商或风电厂等未联合功率可灵活调节资源（例如，储能设备）的新能源发电商现货电力市场单主体决策问题的研究在近年来较为常见，因此，在本文中，“价格接受者”型卖方现货电力市场单主体决策模型的构建是围绕一个风电-储能联合系统（Wind Farm-Energy Storage System, WF-ESS）所展开的。储能设备的快速充放电功能使得联合系统在各平衡市场阶段具备了功率自发调节能力，这使得本文强化学习方法中的 Agent 模型主要是针对一个风电-储能联合系统的 Actor 策略网络和 Critic 价值网络模型。该体系中的所有决策模型均以随机-鲁棒混合优化为数学解析形式。算例仿真验证了本章模型体系的科学性、可行性以及相对于其他现有常见方法的优势。

#### （1）Actor 策略网络模型

Actor 策略网络模型本质上是人工神经网络模型（artificial neural network, ANN），分为输入层、隐含层和输出层。输入层、隐含层和输出层通过权值、阈值和激活函数进行连接，如下图所示。通过数据样本集的训练，权值、阈值将不断迭代调整，使人工神经网络模型构建、学习并存储输入层和输出层之间的映射对应关系。

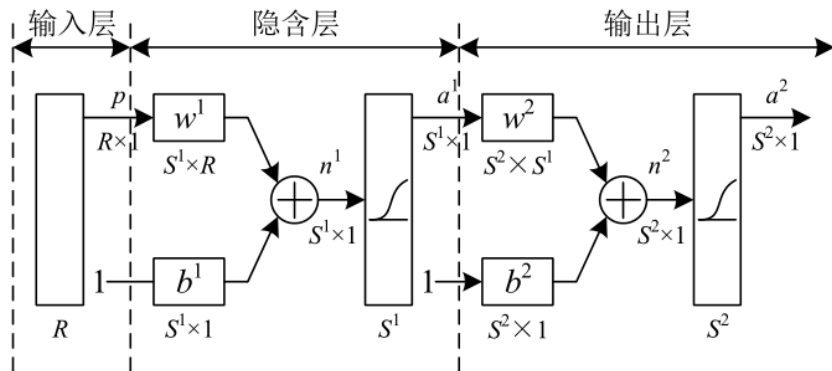


图 4-1 神经网络结构模型

Actor 策略网络的目标是学习并优化策略，使策略的表现越来越好，因此 Actor 策略网络模型的输入为状态特征，输出为动作分布。根据现货电力市场和发电商报价的场景，状态特征包括电力市场状态和发电商自身状态两部分，动作分布为发电商对申报电量和申报电价的决策行为。

#### ① 状态（state）

日前市场交易开展之前，电力交易机构公布该月关键输电通道剩余可用输电容量、该月剩余总用电量需求、该月跨区跨省供应电量、该月各发电商剩余最大发电量等公开市场信息。一方面，发电商根据电力交易机构发布的公开市场信息，判断该月电力市场的供需情况；另一方面，发电商结合自身剩余最大发电量等私有信息，判断自身在该月电力市场中的份额和市场力。因此，Actor 策略网络模型中状态特征包括电力市场状态和发电商自身状态两部分。

a. 电力市场状态

利用“市场供需比”表示电力市场状态，即市场电量供给和市场电量需求的比值。

$$R = \frac{Q_S}{Q_D} = \frac{\sum Q_G^{S,Mon}}{Q_D} \quad (9)$$

式中，

$R$  为市场供需比；

$Q_S$  为市场电量供给；

$Q_D$  为市场电量需求；

$\sum Q_G^{S,Mon}$  储能设备日内最大发电量。

b. 发电商（风电-联合储能系统）自身状态

利用“发电商市场份额”表示发电商自身状态，即发电商剩余最大发电量和市场电量需求的比值。

$$R_s = \frac{\sum Q_G'^{S,Mon}}{Q_D} \quad (10)$$

式中，

$R_s$  为发电商的市场份额；

$Q_D$  为市场电量需求；

$\sum Q_G'^{S,Mon}$  为储能设备日内最大发电量。

因此，Actor 策略网络模型中状态特征可表示为市场供需比和发电商市场份额的二维向量  $[R, R_s]$ 。其中，根据日前现货市场交易的一般情况，认定市场供

需比  $R$  的范围为  $[0.0, 2.0]$ ，发电商市场份额  $R_G$  的范围为  $[0.0, 1.0]$ 。

## ②动作 (action)

发电商根据月度集中竞价交易的市场公共信息和自身私有信息，对申报电量和申报电价进行决策，即为发电商在一定“状态”下所采取的“动作”。发电商在日前现货市场交易中的申报电量和申报电价如下所示。

$$Q_G^{R,Day} = \alpha_G \times Q_G^{S,Day} \quad (11)$$

$$p_G^{R,Day} = \beta_G \times C_G^{M,Day} \quad (12)$$

式中，

$Q_G^{R,Day}$  为发电商在日前现货交易中的申报电量；

$\alpha_G$  为发电商在日前现货交易中的申报电量决策系数；

$Q_G^{S,Day}$  为发电商的剩余最大日发电量；

$p_G^{R,Day}$  为发电商在日前现货市场交易中的申报电价；

$\beta_G$  为发电商在日前现货市场交易中的申报电价决策系数；

$C_G^{M,Day}$  为发电商的日平均边际发电成本。

因此，Actor 策略网络模型中动作分布可表示为申报电量决策系数和申报电价决策系数的二维向量  $[\alpha_G, \beta_G]$ 。其中，发电商在月度集中竞价交易中的申报电量不可能小于零，且不大于其月度剩余最大发电量，认定申报电量决策系数  $\alpha_G$  的范围为  $[0.0, 1.0]$ ；同时一般情况下，发电商在月度集中竞价交易中的申报电价大于等于其月度平均边际发电成本，且根据月度集中竞价交易规则，发电商不得暴利报价，认定申报电价决策系数  $\beta_G$  的范围为  $[1.0, 1.2]$ 。

## (2) Critic 价值网络模型

Critic 价值网络模型本质上也是人工神经网络模型。Critic 价值网络模型的目标是学习并评估值函数，使值函数对策略优劣的评估越来越准确。Critic 价值网络的输入为状态特征，输出为状态值函数。根据中长期电力市场和发电商报价的场景，状态特征包括电力市场状态和发电商自身状态两部分，状态值函数为发电

商在各日前市场交易的收益之和。

Critic 价值网络模型的状态特征与 Actor 策略网络模型的状态特征相同，在此不再赘述。发电商在现货交易的收益即为市场收入和发电成本的差值，其中发电商发电成本是关键因素。

### （3）环境：日前市场交易出清模型

**日前市场出清机制：**设系统中共有 NG 家常规发电商、NW 家新能源发电商、N 个系统节点、NL 条线路，在日前市场中，全天的交易总时段数为 T，日前电力市场基本出清模型如下。

#### 1) 目标函数

在单侧竞价的现货电力市场中，出清模型所追求的目标应为最小化系统运行成本：

$$\min \sum_{t=1}^T \sum_{i=1}^{NG} \rho_{i,t} P_{i,t} + \sum_{t=1}^T \sum_{j=1}^{NW} \rho_{j,t} W_{j,t} \quad (13)$$

式中， $\rho_{i,t}$  和  $\rho_{j,t}$  表示常规发电商 i 和新能源发电商 j 针对时段 t 报出的可接受的最低价格， $P_{i,t}$  和  $W_{j,t}$  分别表示常规及新能源发电商在 t 时段的出力。

#### 2) 节点功率平衡约束

对于系统中的任一节点 n，其流入和流出的有功功率应保持实时平衡，即

$$-L_{n,t} + \sum_{i_n, j_n \in H_n} (P_{i_n,t} + W_{j_n,t}) = \sum_{n' \in \Theta_n} \frac{1}{x_{n,n'}} (\theta_{n,t} - \theta_{n',t}) \quad (14)$$

其中， $H_n$  表示所有连接在节点 n 上的机组及负荷所组成的集合； $\Theta_n$  表示所有与节点 n 上直接相连的节点 n' 所组成的集合。

#### 3) 机组出力约束

在日前出清中，对于常规发电商而言，其各时段机组出力受到其装机容量的限制；对于新能源发电商而言，其各时段机组出力应受到对其实时自然出力预测的限制。表达式如下：

$$0 \leq P_{i,t} \leq P_{i,max}, \forall i, t \quad (15)$$

$$0 \leq W_{j,t} \leq \tilde{W}_{j,t}, \forall j, t \quad (16)$$

式中， $P_{i,max}$  和  $\tilde{W}_{j,t}$  表示常规大电商和新能源发电商的最大容量和最大出力预测值。

#### 4) 机组爬坡约束

常规发电机组由于本身参数的影响及其物理特性的约束,使得机组在两相邻时段内的出力受到如下限制。

$$|P_{i,t} - P_{i,t-1}| \leq ra, \forall i, t \quad (17)$$

式中,  $ra$ 表示机组两相邻时段间的爬坡限制。

#### 5) 线路潮流约束

线路中的有功潮流最大值不能超过线路本身能容纳的最大输电容量,公式表示为

$$\left| \frac{1}{x_{n,n'}} (\theta_{n,t} - \theta_{n',t}) \right| \leq F_{n,n'}, \forall n, t \quad (18)$$

式中,  $F_{n,n'}$ 表示线路能承受的最大传输容量。

### 4.2.4 基于 Actor-Critic 算法的发电商报价模型

市场出清模型的目标函数为所有市场主体的总效用最大化。

市场出清模型的基本原理是“高低匹配,梯级排队”,如下图所示,即需求方(买方)按申报价格从大到小为优先级依次排序,供给方(卖方)按申报价格从小到大为优先级依次排序,模型具体流程如下所示。

(1) 匹配优先级最高(申报价格最大)的需求方和优先级最高(申报价格最小)的供给方,然后匹配优先级次高的需求方和供给方,以此类推;

(2) 当需求方和供给方的申报价格相等,或所有需求方和供给方都完成匹配时,结束出清,完成交易;

(3) 出清价格为供给曲线与需求曲线交叉点对应价格,或最后匹配的需求方和供给方申报价格的平均价格。

假设市场中存在  $n$  个需求方(买方)  $D_n$  和  $m$  个供给方(卖方)  $S_m$ , 其中第  $i$  个需求方(买方)的申报电量和申报电价分别为  $Q_{D_i}$  和  $p_{D_i}$ , 第  $j$  个供给方(卖方)的申报电量和申报电价分别为  $Q_{S_j}$  和  $p_{S_j}$ 。市场出清后,第  $i$  个需求方(买方)和第  $j$  个供给方(卖方)的成交电量和成交电价分别为  $Q_{D_i S_j}$  和  $p_{D_i S_j}$ , 其中成交电价即为市场出清价格(统一边际价格)  $p_{D_i S_j} = p_{MC}$ 。因此,第  $i$  个需求方(买方)和第  $j$  个供给方(卖方)之间的需求方效用为  $U_{D_i}$ , 供给方效用为  $U_{S_j}$ ,

市场总效用为  $U_{D_i S_j}$ 。

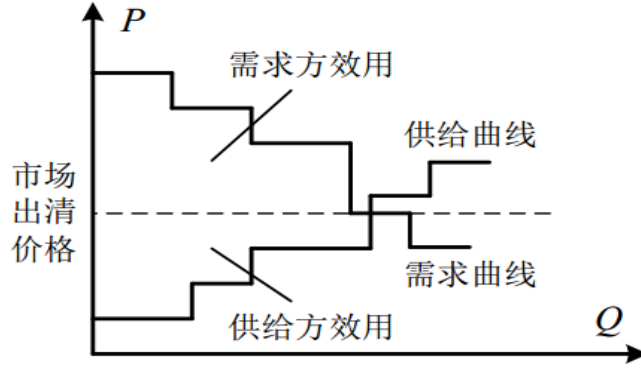


图 4-2 出清原理图

$$U_{D_i} = (p_{D_i} - p_{D_i, S_j}) \times Q_{D_i, S_j} \quad (19)$$

$$U_{S_j} = (p_{D_i, S_j} - p_{S_j}) \times Q_{D_i, S_j} \quad (20)$$

$$U_{D_i, S_j} = U_{D_i} + U_{S_j} = (p_{D_i} - p_{S_j}) \times Q_{D_i, S_j} \quad (21)$$

式中，

$U_{D_i}$  为需求方（买方） $i$  的需求方效用；

$U_{S_j}$  为供给方（卖方） $j$  的供给方效用；

$U_{D_i, S_j}$  为需求方（买方） $i$  和供给方（卖方） $j$  的市场总效用。

因此，在目前中长期电力市场下的市场出清模型如下所示。

$$\text{Maxi} = \sum_{i=1}^n \sum_{j=1}^m U_{D_i, S_j} = \sum_{i=1}^n \sum_{j=1}^m (p_{D_i} - p_{S_j}) \times Q_{D_i, S_j} \quad (22)$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{i=1}^n Q_{D_i, S_j} \leq Q_{S_j} \quad j = 1, 2, \dots, m \\ & \sum_{j=1}^m Q_{D_i, S_j} \leq Q_{D_i} \quad i = 1, 2, \dots, n \\ & Q_{D_i, S_j} \geq 0 \quad i = 1, 2, \dots, n, j = 1, 2, \dots, m \end{aligned}$$

式中，

$p_{D_i}$  为需求方（买方） $i$  的申报电价；

$p_{S_j}$  为供给方（卖方） $j$  的申报电价；

$Q_{D_i, S_j}$  为需求方（买方） $i$  和供给方（卖方） $j$  的成交电量；

$Q_{D_i}$  为需求方（买方） $i$  的申报电量；

$Q_{s_j}$  为供给方（卖方） $j$  的申报电量。

基于 Actor-Critic 算法的发电商报价模型

$$J(\theta) = v_{\pi_\theta}(s_0) + \gamma H[\pi_\theta(a|s)] \quad (23)$$

$$\nabla_\theta J(\theta) = \nabla_\theta v_{\pi_\theta} + \gamma \nabla_\theta H[\pi_\theta(a|s)] \quad (24)$$

式中，

$J(\theta)$  为目标函数；

$\pi_\theta$  为参数化策略；

$\gamma$  为策略交叉熵权重；

$H[\pi_\theta(a|s)]$  为策略交叉熵。

Actor-Critic 强化学习算法的流程如下所示。

（1）输入：状态集合  $S$ 、动作集合  $A$ 、奖励  $R$ 、折现率  $\gamma$ 、策略网络学习率  $\alpha$ 、价值网络学习率  $\beta$ 、策略交叉熵权重  $\gamma$

（2）初始化：

随机设置全局和多线程的策略网络参数  $\theta_G$ 、 $\theta_1$ 、 $\theta_2$ 、...

随机设置全局和多线程的价值网络参数  $w_G$ 、 $w_1$ 、 $w_2$ 、...

（3）策略网络和价值网络参数迭代：

循环：

同步全局策略网络参数： $\theta_1 \leftarrow \theta_G$ 、 $\theta_2 \leftarrow \theta_G$ 、...

同步全局策略网络参数： $w_1 \leftarrow w_G$ 、 $w_2 \leftarrow w_G$ 、...

对于每一线程  $i$ ：

根据策略  $\pi_{\theta_i}(A_i|S_i)$  进行一步实验，返回动作  $a$  和后续状态  $s'$

计算优势函数： $A_i = \hat{q}_{\omega_i}(s, a) - \hat{v}_{\omega_i}(s)$

计算价值网络梯度： $\omega'_i = \omega_i + \beta \nabla_{\omega_i}(A_i)^2$

计算策略网络梯度： $\theta'_i = \theta_i + \alpha \gamma \nabla_{\theta_i} \log \pi_{\theta_i}(a|s) A_i + \gamma \nabla_{\theta_i} H[\pi_{\theta_i}(a|s)]$

异步更新全局策略网络参数： $\theta_G \leftarrow \theta'_i$



异步更新全局价值网络参数:  $w_G \leftarrow w'_i$

直到:  $|\omega'_i - \omega_i| < \sigma, |\theta'_i - \theta_i| < \sigma$

(4) 输出: 最优策略  $\pi_0$ 。

## 4.2.5 平衡市场决策模型构建

### (1) 目标函数

在第  $t(1 \leq \forall t \leq T)$  个平衡市场决策阶段, 联合系统需要对  $P_n^B, W_n, P_n^C, P_n^D (t \leq \forall n \leq T)$  进行决策, 其中,  $P_t^B, W_t, P_t^C, P_t^D$  应根据已经获悉的  $W_t^N$  信息最终确定,  $P_n^B, W_n, P_n^C, P_n^D (t+1 \leq \forall n \leq T)$  的决策结果仅伴随前者共同输出, 但并不真正执行。联合系统在第  $t(1 \leq \forall t \leq T)$  个平衡市场决策阶段进行联合出力调整决策的目标函数为

$$\max_{P_n^B, W_n, P_n^C, P_n^D, E_n} \tilde{R}_{WF-ESS,t} = \sum_{n=t}^T \tilde{\lambda}_n^{B*'} P_n^B - C_W \sum_{n=t}^T W_n - C_E \sum_{n=t}^T E_n \quad (25)$$

式中, 用  $\tilde{R}_{WF-ESS,t}$  表示系统在第  $t(1 \leq \forall t \leq T)$  个平衡市场中的总利润,  $\tilde{\lambda}_n^{B*'}$  表示最近一次更新过的对第  $t$  至  $T$  个平衡市场出清电价的估计。

(2) 系统的约束共包含如下:

#### 1) 功率平衡约束

针对  $n(t \leq \forall n \leq T)$  时段, 系统的申报计划  $P_n^{da*}$ 、 $P_n^B$  与实时功率  $W_n$ 、 $P_n^C$  及  $P_n^D$  之间应存在以下平衡关系

$$P_n^B = (W_n - P_n^C + P_n^D) - P_n^{da*}, t \leq \forall n \leq T \quad (26)$$

对于第  $n$  时段, 其中的  $P_n^{da*}$  为已知量。

#### 2) 风电及储能充放电功率约束

对于联合系统的风机而言,  $t$  时段最终确定的实时出力应受到已经获悉的  $W_t^N$  的限制, 即

$$0 \leq W_t \leq W_t^N \quad (27)$$

同时,  $t$  时段之后的实时出力应受到最近一次更新过的  $t+1$  至  $T$  时段风电实时自然出力估计量  $\tilde{W}_n^{N'} (t+1 \leq \forall n \leq T)$  的限制, 即

$$0 \leq W_n \leq \tilde{W}_n^{N'}, t+1 \leq \forall n \leq T \quad (28)$$

对于储能设备而言，其充放电功率受到自身参数的限制，即

$$0 \leq P_n^C \leq P_{max}^C, t \leq \forall n \leq T \quad (29)$$

$$0 \leq P_n^D \leq P_{max}^D, t \leq \forall n \leq T \quad (30)$$

### 3) 风电机组爬坡约束

由于受机组自身参数的影响，t 时段的实时出力受到 t-1 时段的最终确定的实时出力影响，表达式为

$$|W_t - W_{t-1}^*| \leq rw \quad (31)$$

此外，在 t+1 时段及以后各时段都会存在同样的约束，即

$$|W_n - W_{n-1}| \leq rw, t+1 \leq \forall n \leq T \quad (32)$$

### 4) 储能设备能量平衡约束

对于储能设备，其残余能量应满足以下约束：

$$E_n = E_{n-1}^* + \sum_{i=t}^n P_i^C \eta_C \Delta t - \sum_{i=t}^n (P_i^D / \eta_D) \Delta t, t \leq \forall n \leq T \quad (33)$$

式中， $E_n$ 表示 t 时段储能设备的残余能量， $E_{n-1}^*$ 表示 t-1 时段的储能设备残余能量。

### 5) 储能设备残余能量约束

考虑到储能设备的利用率和寿命等因素，其第 n 时段的残余能量应满足以下约束

$$E_{min} \leq E_t \leq E_{max}, t \leq \forall n \leq T \quad (34)$$

综上，对于整个系统的平衡市场决策模型构建完毕，其输出结果为 $P_t^{B*}$ 、 $W_t^*$ 、 $P_t^{C*}$ 、 $P_t^{D*}$ ，表示在第t( $1 \leq \forall t \leq T$ )个平衡市场决策阶段最终确定的系统出力及储能设备的实时功率。

## 第 5 章 现货电力市场仿真

本文对于一个“价格接受者”型风电-储能联合系统予以展开，对其日前及平衡市场中联合出力、风电实时出力及储能充放电功率进行优化决策。其中相关机组及储能设备的参数如下表。

表 5-1 参数列表

$\eta_c$	$\eta_d$	$E_{min}(\text{MWh})$	$E_{max}(\text{MWh})$	$E_0(\text{MWh})$	$W_{max}(\text{MW})$
0.95	0.95	10	50	30	75
rw(MW)	$P_{max}^C(\text{MW})$	$P_{max}^D(\text{MW})$	$C_E(\text{DK/MWh})$	$C_W(\text{DK/MWh})$	
20	10	10	3.719	196	

用于日前及平衡市场决策的所有出清电价联合随机场景以及风电实时自然出力动态不确定集合均是根据相关历史数据构建并滚动更新而成的。在电价方面，由于我国的现货市场体系尚不完善，因此，日前及平衡市场出清电价历史数据（时点数据）采集于北欧电力市场的 DK-West 地区，采集时间为 2016 年 9 月 1 日至 2016 年 11 月 30 日，由于北欧电力市场中的平衡市场属于双价制结算体系（针对同一个平衡市场阶段上/下调的结算价格之间存在差异，且两者之一与同一交易日的日前市场出清价格相等），因此，我们选取与同一交易日日前市场出清电价存在差异的平衡结算价格作为相应单一价制平衡市场出清电价的历史数据；在风电实时自然出力方面，选取我国河北省张家口市某地区 2016 年 9 月 1 日至 2016 年 11 月 30 日的时段平均风速作为原始数据，并根据文献中的风速-风功率转换曲线进一步将其转化为相应时间段内的平均风电实时自然出力数据。无论是对模型 1 还是模型 2 的仿真，均将 2016 年 11 月 1 日至 30 日当作计算日（即测试日，一方面，根据模型 1、2 对这些日期内联合系统的决策结果进行计算；另一方面，根据电价及风电出力的实际数据对这些日期内联合系统的决策效果进行验证与分析）。对于某一个计算日而言，在该日的前日市场决策阶段，以上所采集的 2016 年 9 月 1 日至该日前一天中午十二点整的时点（日前及平衡）电价及风电实时自然出力数据分别构成相应联合价格随机场景及风电实时自然出力动态不确定集的样本数据集合；在该日的第  $t$  ( $1 \leq t \leq 24$ ) 个平衡市场决策阶段，以上所采集的 2016 年 9 月 1 日至该日  $t-1$  时刻（若此时  $t=1$ ，则  $t-1$  可认为是前一日第 24 个时段）时点（平衡）电价用于更新相应联合价格随机场景的样本数据集合，2016 年 9 月 1 日至该日  $t$  时刻的风电实时自然出力数据用于更新相应风电实时自然出力动态不确定集的样本数据集合。当针对一个计算日的所有日前及平衡市场决策过程执行完毕后，风电-储能联合系统在该日的实际利润是根据当日真实的电价及风电出力数据计算得到的。

在本小节中,风电-储能联合系统在30个计算日内的日前市场联合出力计划,平衡市场联合出力调节计划、实时风电出力与储能充放电功率将基于模型1与模型2分别予以决策。模型1与模型2中的相关参数初步设定为: $\rho=0.5$ ,  $\alpha=0.05$ ,  $\gamma=0.5$ ,  $\text{card}(\Omega_0)=20$ ,  $\text{tcard}(\Omega_t)=20$ ,  $\Delta=0.000001$ 。

图5-1及图5-2以上述30个计算日为例,分别展示了联合系统针对各日在滚动执行了1个模型1及24个模型2后所获得的日前市场联合出力计划,平衡市场联合出力调节计划、实时风电出力与储能充放电功率决策结果。

观察图5-1及图5-2可知:

(1) 在以模型1及模型2的滚动组合为基础进行决策的前提下,一方面,在日前市场中,风电-储能联合系统往往在日前市场价格偏高的时段申报较大的联合出力计划,在日前市场价格偏低的时段申报较小的联合出力计划;另一方面,在平衡市场价格高于日前市场价格的各时段中,风电-储能联合系统往往提供向上的联合出力调节服务,在平衡市场价格低于日前市场价格的各时段中,风电-储能联合系统往往提供向下的联合出力调节服务。上述现象的主要原因在于,具有灵活充放电调节能力的风电-储能联合系统无论在日前市场阶段还是在各平衡市场阶段都具有策略性申报能力。在追求利润最大化的动力以及自身容量、实时风速以及成本等因素的制约下,一方面,日前市场中的“高价多卖”及“低价少卖”策略有助于利润的提升;另一方面,平衡市场中的出清电价反映了系统对功率调节方向的需求情况,根据第2章中的相关理论可知,在单一价结算体系下,若平衡市场出清电价高于相应时段的日前电价,则表明市场总体存在实时出力“短缺”(总出力低于总负荷)问题,提供向上的出力调节能使得联合系统进一步获益,反之,若平衡市场出清电价低于相应时段的日前电价,则表明市场总体存在实时出力“过剩”(总出力高于总负荷)问题,提供向下的出力调节能使得联合系统进一步获益。然而,出清电价及风电实时自然出力的不确定性有时会干扰联合系统对联合出力申报策略的决策,这就是图3-2中,第14、530等时段(分别对应于11月1日14时(时段)及11月22日2时(时段))的日前联合出力计划,第275、369等时段(分别对应于11月11日11时(时段)及11月15日9时(时段))的平衡市场联合出力调节计划与上述规律相背离的主要原因。

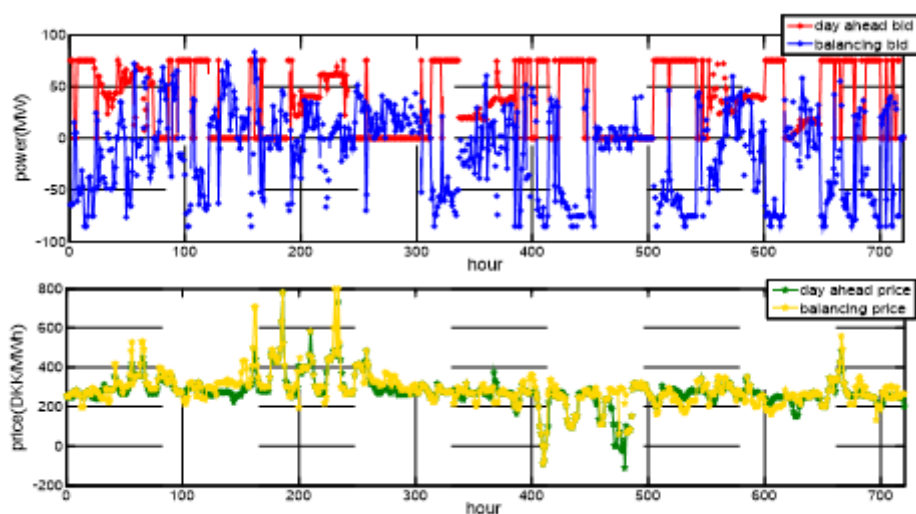


图 5-1 风电-储能联合系统的日前市场联合出力计划及平衡市场联合出力调整计划

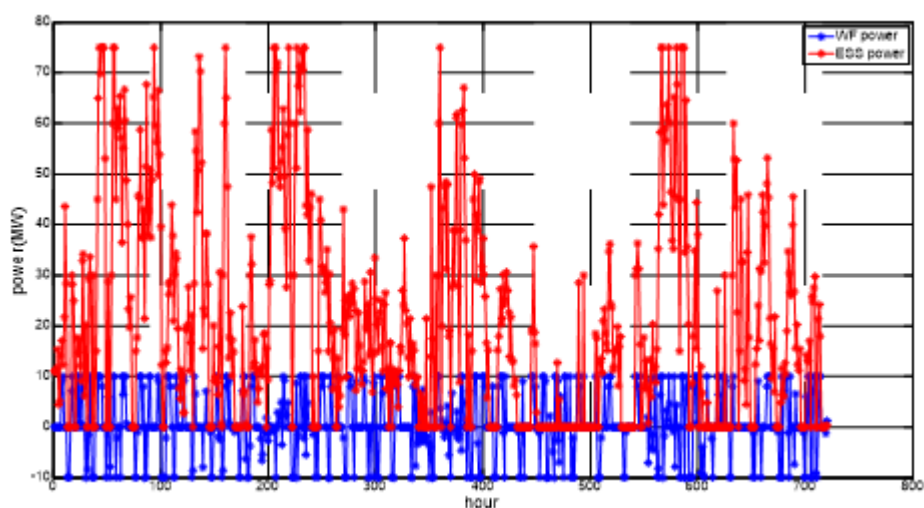


图 5-2 联合系统实时风力出力与储能充放电功率

(2) 风电-储能联合系统在任意时段的风电实时出力以及储能充/放电功率在各自满足相应的风机及储能设备功率及能量约束外,与同时段的日前市场联合出力计划、平衡市场联合出力调节计划之间满足如下功率平衡关系:  $P_t^B = (W_t - P_t^C + P_t^D) - P_t^{da}, 1 \leq \forall t \leq 24$ 。这意味着在以模型 1 及模型 2 的滚动组合为基础进行决策的前提下,风电-储能联合系统总能够通过合理调度自身风机与储能设备的有功功率以实现联合系统在日前及平衡市场中的策略性申报结果。另外,联合图 5-2 及 5-3 还不难看出,储能设备往往是在平衡电价较高的时段选择放电,在平衡电价较低的时段选择充电,这种“高价放电”及“低价充电”充放电策略

也是进一步提升联合系统利润的必要手段,是本文模型所输出的关键性决策结果之一。然而,与上文分析相类似,出清电价及风电实时自然出力的不确定性有时也会干扰联合系统对储能充放电策略的决策,这就是图 3-6 中,第 200、405 等时段(分别对应于 11 月 8 日 8 时(时段)及 11 月 16 日 16 时(时段))的储能充/放电功率与上述规律相背离的主要原因。

综上所述,在以模型 1 及模型 2 的滚动组合为基础进行决策的前提下,一方面能使联合系统(作为一个整体)同时在日前及平衡市场中进行策略性联合出力申报;另一方面可调度储能设备实现“高价放电”及“低价充电”充放电策略,即实现了储能设备在市场中的套利功能。当针对 30 个计算日的所有日前及平衡市场决策过程执行完毕后,联合系统在各日所获得实际利润如表 5-1 所示:

表 5-1 风电-储能联合系统在 30 个计算日内所获得的实际利润

日期	Nov.2016									
	1	2	3	4	5	6	7	8	9	10
利润		126279.	194830.	130302.	93546.3	59157.3	98034.1	111969.	252915.	267336.
(DKK)	45258.2	3	3	9	7	4	8	1	1	5
日期	11	12	13	14	15	16	17	18	19	20
利润	133869.	40408.9	43821.8	48287.7	62051.2	124755.	47639.8	34142.7	24995.5	9879.36
(DKK)	1	3	2	5	5	8	4	3	7	
日期	21	22	23	24	25	26	27	28	29	30
利润	53345.5	74950.4	78977.2	95204.3	125490.	69980.9	37642.4	114004.	48217.8	22573.7
(DKK)	1	7	7	9	8	4	8	2	4	7

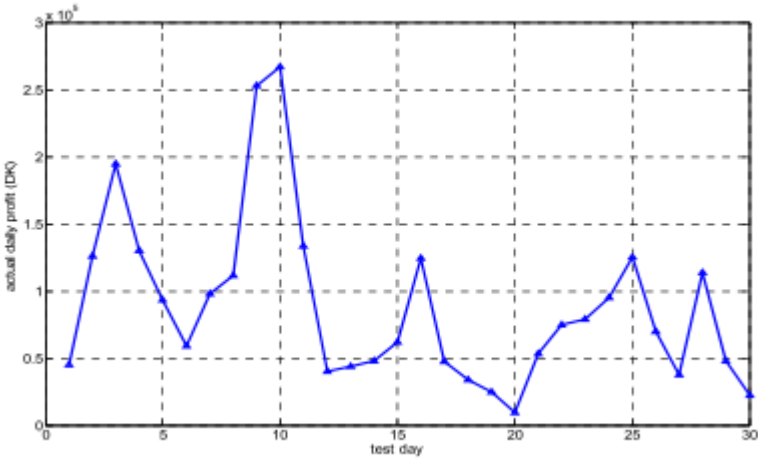


图 5-3 风电-储能联合系统在 30 个计算日内所获得的实际利润

另外,需要说明的是,通过在上述 30 个计算日内的仿真还可发现,基于模型 1 执行日前市场决策的平均计算时间约为 9.7 秒,基于模型 2 执行平衡市场决策的平均计算时间约为 5.2 秒。然而,根据第二章中的相关理论可知,日前市场

通常至少提前 12 个小时开展，平衡市场至少提前数分钟开展，由此可见，本章所构建的模型 1 及模型 2 的计算时间均远小于相应市场阶段的提前时间量，用其进行相应市场阶段的决策在计算时间上是完全可行。

本小节中，为了简化描述，将风电-储能联合系统基于模型 1 及模型 2 所生成的日前、平衡市场联合申报计划以及风机、储能设备的实时出力及功率决策结果称为策略 1。一般情况下，风电-储能联合系统日前及平衡市场决策模型是基于随机优化及线性仿射方法所构建的，其中，日前市场决策模型基于随机优化方法所构建、平衡市场决策模型是基于线性仿射方法所构建（具体的模型构建方法请详见文献[23]）。此处将风电-储能联合系统基于文献[23]模型所生成的日前市场联合申报计划以及风机、储能设备的实时出力及功率决策结果称为策略 2。策略 2 相较于其他决策方法在决策效果等方面的优势已经在文献[23]中得到了充分的验证，因此，本小节的主要任务是，在相同的基础历史数据、模型参数取值以及随机场景生成方法条件下，通过比较策略 1 及策略 2 在上述 30 个计算日内所得到的实际利润，以验证模型 1 及模型 2 的优势。

下图分别展示了策略 1 与策略 2 在 30 个计算日内所获得的实际利润：

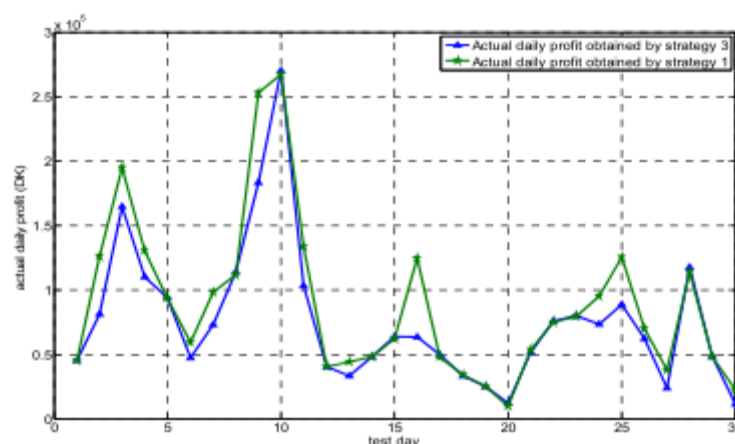


图 5-4 策略 1 及策略 2 在 30 个计算日内所获得的实际利润比较

从图 5-1 中可以看出，在大多数计算日中，风电-储能联合系统通过策略 1 所获得的实际利润要高于策略 2。这基本验证了本章所构建的模型 1 及模型 2（相较于文献[23]中所构建与提及的所有模型）在辅助联合系统现货市场决策获利方面的显著优势。本文认为，其主要原因在于：

（1）文献[23]所构建的基于线性仿射的平衡市场决策模型，在一定程度上限

制了实时风电出力及储能充/放电功率调整的可行域范围，这是因为这种用于平衡阶段决策的线性仿射模型（模型参数）实质上仅是由日前决策模型所决策生成的，而在随后的次日各平衡市场决策阶段并未因新获取的电价及风电自然出力信息而进行调整（该线性仿射模型反映了电价及风电自然出力偏差（实时量和日前预测值之间的偏差）与风电实时出力、储能充/放电功率决策值之间的映射关系，详见文献[23]）。然而，在本章的模型体系中，模型 1 并不对模型 2 的构建形式及主要参、系数取值进行决策，模型 2 对实时风电出力及储能充/放电功率的（再）决策是基于实时更新的电价及风电自然出力信息而动态进行的。这样做的好处在于，一方面，不以日前决策结果限制平衡决策模型可提高平衡市场阶段决策的灵活性，降低日前决策结果对平衡决策可行域范围的干扰，提高日前及平衡决策的整体效率；另一方面，基于实时更新的相关电价及风电实时自然出力信息，可最大限度提高不确定因素预测的精准性，缩小其随机波动范围，降低模型决策的保守性。

（2）文献[23]忽略了风电-储能联合系统在平衡市场中进行策略性联合出力申报的可能性，即并未在各平衡市场决策阶段对联合出力调整计划申报量进行决策。

## 第 6 章 结论与展望

### 6.1 结论

现货电力市场是电力市场体系中的重要环节，其发现价格、平衡功率、维持系统安全稳定运行等方面的重要作用使之越来越成为一国或一个地区电力市场体系建设相对完善的重要标志。现货市场是由交易组织者及各交易参与者等经济行为决策主体所共同组成的系统，整个市场会在所有主体共同合理决策的作用下实现对电力资源的优化配置。由此可见，无论是从交易参与或组织者角度，还是从全市场的角度，合理进行或引导各类经济决策均是保障现货电力市场有效运行的重要手段。

本报告在充分调研了广东省现货开展情况的背景下，考虑新能源发电商参与现货市场交易的基础上，从单主体视角出发，构建了基于不确定优化以及 Actor-



Critic 强化学习算法的现货电力市场发电商报价模型，并基于 Agent 对现货电力市场进行仿真，主要研究结论如下。

### 6.1.1 决策模型小结

本文建立的单主体决策模型体系能够辅助“价格接受者”型风电-储能联合系统在复杂的现货市场环境下进行联合出力计划及其调整决策，从而，一方面，确保储能设备平抑功率偏差与促进策略性申报等功能的合理实现；另一方面，通过充分结合随机及鲁棒优化方法各自的特点及优势，提高了决策的优化效果。另外，前文关于模型在计算时间等方面的展示客观上决定了本章模型体系较易在实际应用中进行推广。

### 6.1.2 电力市场监管建议

(1) 现货电力市场中，相比于经济持留行为，电力交易中心应更侧重于对发电商物理持留行为的监管。发电商发电成本属于非公开的私有信息，电力交易中心对发电商申报电价的监管往往具有较大操作和实施难度。但是，装机规模属于市场公开信息，因此电力交易中心通过对比发电商申报电量与其装机规模和剩余发电量，能够较轻易发现发电商物理持留行为。

(2) 不同市场力、不同发电成本的发电商采取物理持留行为，对市场出清电价和市场效率的影响不同。由于市场力较大、发电成本较低的发电商采取物理持留行为，将提高市场出清电价，降低市场效率，因此电力交易中心应更侧重于该类型发电商的监管和惩罚。

## 6.2 展望

本论文基于不确定优化以及强化学习算法等技术，对现货市场发电商报价策略进行研究，但是还有以下两个方面有待改进。

(1) 本论文在强化学习 Agent 的 Actor 策略网络模型中，模型输入状态特征仅考虑电力交易中心披露的公共市场状态和发电商的自身状态。但是，电力市场是一种具有较强寡头垄断特性的市场，发电商之间具有一定概率采取串谋的报

价行为。因此，在 Actor 策略网络模型中考虑发电商串谋的私有非对称信息，构建基于强化学习算法的对称/不对称信息下电力市场发电商报价模型和市场仿真技术，是后续研究方向之一。

（2）现货电力市场中市场主体除了发电商，还有售电公司以及大用户。本文考虑的市场主体比较单一，尚未构建多主体智能决策模型。随着电力市场的逐步成熟，需求侧在电力市场中扮演的角色将会越来越重要，构建多主体博弈的强化学习决策模型，是另一后续研究方向。

## 参考文献

- [1] 伏开宝.电力体制改革中市场价格波动管理研究[D].上海社会科学院,2018.
- [2] 李博,高志远.人工智能技术在智能电网中的应用分析和展望[J].中国电力,2017,50(12):136-140.
- [3] 余涛,周斌,甄卫国.强化学习理论在电力系统中的应用及展望[J].电力系统保护与控制,2009,37(14):122-128.
- [4] 《广东省“互联网+现代水利”行动计划》印发实施[J].广东水利水电,2017(03):70.
- [5] 董锦瑜.基于企业经营决策的发电商最优竞价策略[J].经贸实践,2016(09):80.
- [6] 高鑫,王秀丽,雷兵,王锡凡.独立发电商的策略报价研究[J].中国电机工程学报,2004(07):45-50.
- [7] 王云凯,蓝金辉.ML-FFA:基于机器学习和基本面因子分析的量化投资策略[J].时代金融,2018(32):358-359+375.
- [8] 何建华,何祖威,胡效雷.基于成本分析的火电厂日前市场竞价上网报价[J].重庆大学学报,2008(03):276-279+323.
- [9] 宋嗣博,杨苹,许志荣,黄梓健,张育嘉,喻磊,黄锦成.基于日前电价预测的机组报价策略[J].南方电网技术,2017,11(02):57-62+70.
- [10] 刘建良,周杰娜,杨华.基于双人博弈且参数估计下的发电商报价策略研究[J].中国电机工程学报,2007(19):62-67.
- [11] 李清清,周建中,莫莉,张勇传.基于序贯博弈模型的发电商竞价策略分析[J].电

网技术,2008(14):85-89.

- [12]张海峰.考虑风电的电力市场发电商竞标博弈模型[J].电子测量与仪器学报,2016,30(11):1701-1708.
- [13]Maryam Hajati,Hossein Seifi,Mohamad Kazem Sheikh-El-Eslami. Optimal retailer bidding in a DA market – a new method considering risk and demand elasticity[J]. Energy,2010,36(2).
- [14]杨萌. 电力市场下计及风险规避的售电公司购售电策略研究[D].华北电力大学(北京),2017.
- [15]Kazem Zare,Antonio J. Conejo,Miguel Carrión,Mohsen Parsa Moghaddam. Multi-market energy procurement for a large consumer using a risk-aversion procedure[J]. Electric Power Systems Research,2009,80(1).
- [16]卢强,陈来军,梅生伟.博弈论在电力系统中典型应用及若干展望[J].中国电机工程学报,2014,34(29):5009-5017.
- [17]Sayyad Nojavan,Behnam Mohammadi-Ivatloo,Kazem Zare. Robust optimization based price-taker retailer bidding strategy under pool market price uncertainty[J]. International Journal of Electrical Power and Energy Systems,2015,73.
- [18]R.G. Karandikar,S.A. Khaparde,S.V. Kulkarni. Strategic evaluation of bilateral contract for electricity retailer in restructured power market[J]. International Journal of Electrical Power and Energy Systems,2009,32(5).
- [19]Sayyad Nojavan,Hadi Ghesmati,Kazem Zare. Robust optimal offering strategy of large consumer using IGDT considering demand response programs[J]. Electric Power Systems Research,2016,130.
- [20]赵会茹,赵名锐,王玉玮.售电侧开放下日前电力市场动态博弈模型[J].电力建设,2017,38(04):144-152.
- [21]肖云鹏,王锡凡,王秀丽,别朝红.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(03):663-674.
- [22]Alberto Borghetti,Stefano Massucco,Federico Silvestro. Influence of feasibility constraints on the bidding strategy selection in a day-ahead electricity market session[J]. Electric Power Systems Research,2009,79(12).

- [23] Ding H, Pinson P, Hu Z, et al. Optimal offering and operating strategies for wind-storage systems with linear decision rules[J]. IEEE Transactions on Power Systems, 2016, 31(6): 4755-4764.
- [24] Zhen Liu, Jianming Yan, Yuren Shi, Kaiwei Zhu, Gangqing Pu. Multi-agent based experimental analysis on bidding mechanism in electricity auction markets[J]. International Journal of Electrical Power and Energy Systems, 2012, 43(1).
- [25] J. Vijaya Kumar, D.M. Vinod Kumar, K. Edukondalu. Strategic bidding using fuzzy adaptive gravitational search algorithm in a pool based electricity market[J]. Applied Soft Computing Journal, 2013, 13(5).
- [26] Wietze Lise, Vincent Linderhof, Onno Kuik, Claudia Kemfert, Robert Östling, Thomas Heinzow. A game theoretic model of the Northwestern European electricity market—market power and the environment[J]. Energy Policy, 2005, 34(15).
- [27] Erling Pettersen, Andrew B. Philpott, Stein W. Wallace. An electricity market game between consumers, retailers and network operators[J]. Decision Support Systems, 2004, 40(3).
- [28] 程可. 基于多智能体的电力市场竞价行为研究[D]. 华北电力大学, 2015.
- [29] BACH THANH QUY (白青贵). 电力批发市场中基于强化学习的参与者行为特性研究[D]. 湖南大学, 2013.
- [30] 王海宁. 基于复杂系统多 Agent 建模的电力市场仿真技术研究[D]. 中国电力科学研究院, 2013.
- [31] E. Ela, M. Milligan, A. Bloom, A. Botterud, A. Townsend, T. Levin, B.A. Frew. Wholesale electricity market design with increasing levels of renewable generation: Incentivizing flexibility in system operations[J]. The Electricity Journal, 2016, 29(4).
- [32] Mojtaba Shivaie, Mohammad T. Ameli. An environmental/techno-economic approach for bidding strategy in security-constrained electricity markets by a bi-level harmony search algorithm[J]. Renewable Energy, 2015, 83.
- [33] 许子智, 曾鸣. 美国电力市场发展分析及对我国电力市场建设的启示[J]. 电网技

术,2011,35(06):161-166.

- [34] 曾嘉志,赵雄飞,李静,李刚,李康敬,魏震波.用电侧市场放开下的电力市场多主体博弈[J].电力系统自动化,2017,41(24):129-136.