

分类号: TM73

密 级: 公开

学校代码: 10363

学 号: 2210330171



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题 目 基于深度强化学习的微电网
能源调度研究

论 文 作 者	宫家凯
校 内 导 师	余诺
校 外 导 师	赵义正
专业学位类别	电子信息 (085400)
研究方向 (领域)	人工智能

论文提交日期: 2024 年 5 月 31 日

分类号: TM73
密 级: 公开

单位代码: 10363
学 号: 2210330171

题 目 基于深度强化学习的微电网能源调度研究

英文并列

题 目 Research on Energy Dispatching of Microgrid Based
on Deep Reinforcement Learning

学生姓名: 宫家凯

校内导师: 余 诺

校外导师: 赵义正

专业学位类别: 电子信息(085400)

研究方向(领域): 人工智能

论文答辩日期: 2024 年 5 月 19 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律结果由本人承担。

学位论文作者签名：曾家凯

日期：2024 年 05 月 24 日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在 ____ 年解密后适用本版权书。

本学位论文属于

不保密 ☒。

学位论文作者签名：李凯

日期：2024 年 05 月 24 日

指导教师签名：余诺

日期：2024 年 05 月 24 日

基于深度强化学习的微电网能源调度研究

摘 要

随着社会的发展，化石能源的消耗日益增加，导致化石能源供应量减少和其产生的环境污染量增加。微电网作为新型电力系统重要组成部分，可以通过对能源的协调控制来提高能源的使用效率，缓解化石能源的消耗以及污染气体的排放。然而，微电网系统规模的不断扩大和供电侧间歇性、波动性的特点，给微电网能源调度和管理带来了诸多挑战。鉴于深度强化学习可以处理不确定环境下的序列决策问题，本文基于深度强化学习对微电网系统能源调度方法进行深入研究，建立微电网储能电池充放电调度策略、构建电池储能荷电状态信息缺失下多微电网能源调度机制、并提出了多目标综合能源协同优化策略。本文的主要内容包括以下三个方面：

针对基于深度强化学习微电网系统调度优化下的行为价值评估偏差的问题，通过阐述电池储能的运行机理，分析电池储能设备对微电网系统运行产生的作用，考虑电价及可再生能源发电在时间上的波动、随机性特点。进而提出一种基于改进的深度 Q 网络方法，建立深度强化学习的电池充放电调度决策，对微电网运行状态下储能调度的行为价值实现了精准的评估。

针对多微电网运行过程中电池储能荷电状态信息缺失，加剧系统最优能源调度困难的问题，建立了多微电网协作运行机制，构建多微

电网系统的多智能体马尔可夫决策过程，通过提出的递归多智能体强化学习算法中的长短期记忆神经网络，预测系统多微电网储能荷电状态信息，实现微电网系统非完全观测状态下经济运行的调度优化。

针对综合能源微电网系统多目标调度优化中不同目标相互制约，导致难以优化各自目标权重下多目标的问题，分析转移负荷、消减负荷需求响应机制，建立用户用电、热水、室内温度的综合不满意度模型。基于实时电价、可再生能源发电、室外温度和用户热水需求不确定性及不同类型用能设备运行特点多样性，提出了多目标的微电网综合能源调度优化的方法，实现了用户不满意度下降的同时系统经济低碳运行的优化策略。

关键词：微电网，调度优化，强化学习，多智能体，多目标

Research on Energy Dispatching of Microgrid Based on Deep Reinforcement Learning

ABSTRACT

With the development of society, the consumption of fossil energy is increasing day by day, which leads to the decrease of fossil energy supply and the increase of environmental pollution. Microgrid, as an important part of the new power system, can improve the efficiency of energy use through coordinated control of energy, and alleviate the consumption of fossil energy and the emission of polluting gases. However, the continuous expansion of the scale of microgrid system and the intermittent and fluctuating characteristics of the power supply side have brought many challenges to the energy dispatching and management of microgrid. In view of deep reinforcement learning can deal with sequential decision-making problems in uncertain environment, this paper studies the energy scheduling method of microgrid system based on deep reinforcement learning. On the basis of establishing the charging and discharging optimization model of energy storage batteries in microgrid, the energy scheduling mechanism of microgrid under the lack of energy storage status information of multi-batteries is constructed, and a multi-objective comprehensive energy collaborative optimization strategy is proposed. The main contents of this paper include the following three aspects:

Aiming at the problem of behavior value evaluation deviation based on optimal dispatching of microgrid system, this paper expounds the operation mechanism of battery energy storage, analyzes the effect of battery energy storage equipment on the operation of microgrid system, and considers the fluctuation and randomness characteristics of electricity

price and renewable energy generation in time. Then, an improved deep Q network method is proposed, and a battery charging and discharging decision model based on deep reinforcement learning is established, which accurately evaluates the behavior value of energy storage scheduling in microgrid operation.

In view of the lack of battery energy storage charge state information during the operation of multi-microgrid, which aggravates the difficulty of optimal energy dispatch in the system, a multi-microgrid cooperative operation mechanism is established, and a multi-agent Markov decision process for the multi-microgrid system is constructed. Through the long-short-term memory neural network in the proposed recursive multi-agent reinforcement learning algorithm, the state information of energy storage in the system is predicted, and the dispatch optimization of the economic operation of the microgrid system in the incomplete observation state is realized.

Aiming at the problem that different objectives restrict each other in multi-objective optimal dispatching of integrated energy microgrid system, which makes it difficult to optimize multi-objectives under their corresponding weights, the response mechanism of load transfer and load reduction demand is analyzed, and the comprehensive dissatisfaction model of users' electricity consumption, hot water and indoor temperature is established. Based on the uncertainty of real-time electricity price, renewable energy generation, outdoor temperature and hot water demand of users and the diversity of operation characteristics of different types of energy-using equipment, a multi-objective integrated energy optimal dispatching method for microgrid is proposed, which realizes the optimization strategy of decreasing user dissatisfaction and economic low-carbon operation of the system.

Jiakai Gong (Original draft)

Supervised by Prof. Nuo Yu

KEY WORDS: microgrid, optimal dispatching, reinforcement learning,
multi-agent, multi-objective

目 录

第 1 章 绪论	1
1.1 研究背景及意义	1
1.2 国内外发展现状	2
1.2.1 微电网调度优化模型研究现状	2
1.2.2 微电网调度优化方法研究现状	3
1.2.3 强化学习算法研究现状	4
1.3 主要研究内容	5
1.3.1 微电网储能设备充放电调度	5
1.3.2 多微电网协同能源调度	6
1.3.3 综合能源微电网多目标调度	6
1.4 论文的组织结构	7
第 2 章 强化学习相关理论及算法	9
2.1 强化学习基础理论	9
2.1.1 强化学习背景	9
2.1.2 马尔可夫决策过程	9
2.2 强化学习基础算法原理	11
2.2.1 动态规划算法	11
2.2.2 蒙特卡洛算法	11
2.2.3 时序差分算法	12
2.3 本章小结	12
第 3 章 基于深度强化学习的微电网储能充放电调度优化	13
3.1 引言	13
3.1 系统模型描述与问题建模	13
3.1.1 系统模型描述	13
3.1.2 问题建模	14
3.2 微电网储能充放电调度优化算法设计	15
3.2.1 马尔可夫决策过程构建	15
3.2.2 最大均值差异深度 Q 网络	16
3.3 仿真实验与结果分析	17
3.3.1 仿真实验配置	17
3.3.2 性能分析	18
3.3.3 系统运行成本对比结果	19
3.4 本章小结	21
第 4 章 基于深度强化学习的多微电网协同能源调度优化	22
4.1 引言	22
4.1 系统模型描述与问题建模	22
4.1.1 系统模型描述	22
4.1.2 协作运行的多微电网系统问题建模	25
4.2 多微电网协同能源调度优化算法设计	25
4.2.1 部分可观测马尔可夫决策过程	25
4.2.2 多智能体强化学习算法	26

4.3 仿真实验与结果分析	28
4.3.1 仿真实验配置	28
4.3.2 算法性能对比	30
4.3.3 仿真结果分析	30
4.4 本章小结	31
第 5 章 基于深度强化学习的综合能源微电网能源调度优化	32
5.1 引言	32
5.1 系统模型描述与问题建模	32
5.1.1 系统模型描述	32
5.1.2 问题建模	36
5.2 综合能源微电网多目标调度优化算法设计	38
5.2.1 多目标马尔可夫决策过程	38
5.2.2 多目标强化学习算法	40
5.3 仿真实验与结果分析	42
5.3.1 仿真参数	42
5.3.2 性能对比	46
5.3.3 仿真结果	47
5.4 本章小结	51
第 6 章 总结与展望	52
6.1 总结	52
6.2 未来展望	53
参考文献	54
攻读学位期间发表的学术成果	60
致 谢	61

第 1 章 绪论

1.1 研究背景及意义

工业和居民用电随着国民经济的发展而日益增长,能源需求和用电形式越来越呈现出多样化、个性化的发展趋势,用户对供电的可靠性和供电需求的灵活性要求越来越高,传统大电网逐渐难以满足^[1,2]。构建清洁、低碳、安全、高效的新型电力系统成为目前社会的迫切需求。在“双碳”的大背景下,太阳能、风力发电等可再生能源占比逐渐提高,发展规模迅速扩大,但可再生能源具有随机性和波动性的不可控性特点,大规模应用及接入也给传统电网带来巨大的挑战及冲击^[3],需要提升可再生能源就地消纳水平及电力系统自平衡能力。

为了应对可再生能源大规模接入电网带来的冲击,微电网成为学术界和工业界广泛的关注对象,并且成为各国政府倡导的新能源发展路径^[4,5]。微电网主要包括分布式电源、储能设备、用电负荷、能量转换装置、监控系统和保护装置等主体,是涵盖“发-输-配-用”等环节的自治型电力系统,可以缓解可再生能源对电力系统稳定运行的影响,对实现智微电网的调度优化问题有重要意义。本文针对微电网系统,充分优化储能设备、分布式电源、需求侧灵活负荷的协同调节能力,以应对可再生能源的接入与外部电价信号的波动对微电网系统运行的影响,实现智能微电网的经济低碳运行。

微电网运行调度优化问题是一个序列决策问题,采用传统的方法会消耗大量的计算资源与时间,同时,对于外部环境引起模型数据变动的情况,需要重新建立精确的模型并且调整各项参数^[6]。深度强化学习作为一种数据驱动的方法,能够减少对物理建模的依赖,通过神经网络拟合系统数值,高效处理多维度的复杂信息,从而形成具有很强兼容性的自适应优化决策方法^[7-9]。因此,基于深度神经网络的数据驱动方式逐渐体现了优越的控制优势,尤其是深度强化学习在微电网中的应用被众多学者开展了广泛的研究。但在目前的研究中,深度强化学习在微电网调度优化中仍然有一些问题需要探索:

1) 使用强化学习的方法构建微电网系统调度优化策略,有效降低求解复杂系统调度优化模型的难度,已有很多学者提出基于深度强化学习微电网系统调度策略,其中神经网络对微电网系统调度策略的动作价值精准评估可以减少调度优化的误差。但在目前研究过程中,缺乏对基于深度强化学习微电网储能调度优化

策略的动作价值评估的探索,不能完全体现储能充放电对微电网经济运行的作用,因此亟需开展基于强化学习精确评估动作价值的微电网储能调度研究。

2) 在微电网优化调节过程中,需要考虑多微电网同时协作运行的情况,在多微电网协调运行过程中需要彼此储能荷电状态信息的交互。而用户出于保护用电隐私或者通信受阻致使荷电状态信息缺失,都将增加深度强化学习优化的困难。此时需要进行多微电网能源协同调度的同时预测储能荷电状态。因此,如何在非完全观测状态下构建多微电网系统能源调度机制有着重要的意义。

3) 考虑到综合能源微电网中不同种类的电源和设备,如太阳能、风能、热能、冷能及化学储能等,以及用户侧参与需求响应的场景,多种能源的综合利用以及灵活需求侧的优化使得微电网运行的优化问题变得更加复杂,需要考虑多个变量和约束条件,是具有非线性的组合优化问题。另外在实际微电网运行中会同时考虑多个目标,如用户不满意度、经济性、环保性等。如何实现综合能源微电网系统多目标能源调度优化是目前需要进一步研究的。

1.2 国内外发展现状

随着微电网技术的不断发展,微电网能源调度逐渐成为了研究热点,微电网能源调度是对微电网中各分布式发电设备、储能设备、负荷等元件的能量需求进行合理的调度,以实现微电网系统运行成本、消耗功率、环境污染排放量、用户用电舒适度等目标的优化。进行微电网能源调度优化需要对微电网内各单元建立数学模型,然后构建系统运行的目标函数以及对设备单元进行约束,最后采取优化算法进行能源调度优化策略的求解。

1.2.1 微电网调度优化模型研究现状

微电网调度模型优化主要是通过控制分布式发电单元的输出功率、储能设备的充放电功率、负荷需求量的削减与转移等控制变量的优化。调度模型包括目标函数和约束条件两部分。

微电网能源调度的优化目标一般考虑电网的经济性、环保性、用电质量等等。文献[10]构建了大规模分布式能源的微电网模型,对微电网系统运行成本进行调度优化,实现系统运行的经济性。文献[11]提出了电池储能和热储能结合的微电网系统,对电池储能和热储能的容量配置进行优化,降低混合储能的能源调度成

本。文献[12]在风电发电不确定下提出一种计及数据中心的微电网经济调度模型,可以有效降低系统运行成本和灵活的调整风力出力的不确定性。文献[13]建立系统总运行成本最小为目标的微电网调度模型,同时引入环保性参数,实现微电网的经济性与环保性互补发电调度。文献[14,15]提出了将二氧化碳气体排放最小化为优化目标,并结合可再生能源提高系统能源的经济性,实现对系统经济低碳的运行目的。文献[16]建立计及可中断负荷的多目标调度策略数学模型,目的是实现微电网最小化运行成本,同时减少联络线波动对上级电网的影响。文献[17]描述了一种最优调度技术,该技术实现微电网系统在一天时间段内运行的最小污染排放和能源消耗。

微电网能源调度的约束条件是微电网安全稳定运行的保证,在实现微电网调度优化过程中需要满足一定的约束条件。考虑微电网系统的网络层方面需要满足潮流约束,包括各节点的运行电压、有功功率、无功功率、相角约束^[18];微电网与外部电网联络线传输功率约束^[19];考虑微电网系统的单元层方面需要满足各单元出力的上下限、荷电状态约束、储能系统容量约束等等^[20]。

1.2.2 微电网调度优化方法研究现状

目前,随着微电网成为研究的热点,许多学者已经对微电网调度优化方法进行了研究,并实现了较好的效果。其主要研究方法包括经典数学优化方法、启发式算法、随机优化方法、鲁棒优化方法等等。

对于微电网系统能量调度优化问题。文献[21]遗传算法解决储能设备的容量配置和充电/放电优化问题,目的是降低系统运行成本的总和。文献[22]将调度问题建模为随机混合整数线性规划问题,研究最优随机储能调度,并评估混合储能系统在并网模式下微电网的预期性能。文献[23]考虑可再生能源发电的随机性,建立了储能设备运行方案的日前调度策略,并提出了一种机会约束规划方法来降低系统的运行成本。文献[24]利用随机模型预测控制技术来解决储能运行控制问题,其考虑了在调度区间可能存在并网和孤岛模式之间的转换。文献[25-29]采用了多种优化方法,如启发式算法,以提高微电网运行的经济性。然而,上述方法大多依赖精确的微电网系统数学模型,在实际应用中需要大量计算资源,存在求得的优化解易陷入局部最优的问题。

对于综合能源微电网能源调度优化问题,文献[30]提出了一种混合整数线性规划方法来提高综合能源系统运营商和需求响应聚合商的利润的同时实现削峰

填谷,减少对电网安全稳定运行造成的影响。文献[31]介绍了解决电-气-体化系统最优调度模型问题的混合整数线性规划方法,该方法旨在降低运行成本和碳排放。文献[32]中采用了交替方向乘法器方法来适应多能量互补微电网中的可再生能源。文献[33]提出了分布式能源和微电网运营商聚合器的交易能量共享方法,以最大限度地提高系统的总运行成本。虽然上述算法在解决最优调度问题方面有了显著的改进,但对于实际的综合能源微电网优化问题,该算法仍然依赖于隐式枚举,导致计算负担大,系统复杂性增加^[34]。

1.2.3 强化学习算法研究现状

近年来,在人工智能技术的广泛应用下,深度强化学习(Deep Reinforcement Learning, DRL)为解决微电网系统能量调度问题提供了新的思路。DRL 是一种数据驱动的方法,它具有较好的自适应学习能力和非凸非线性问题优化决策能力,避免了像传统方法简化模型的非凸和非线性部分来解决问题,同时, DRL 可以通过智能体与环境的交互进行学习能源调度优化策略,并不需要精确的数学模型。

随着人工智能技术的发展,深度强化学习方法已开始被用于优化能量存储和调度^[35-37]。文献[38]提出了一种 Q 学习(Q-learning)方法来管理储能系统的电力调度计划,用于解决电价不确定性造成难以实现调度优化的问题。文献[39]将蓄电池和储氢设备的动作(充电、放电和不工作)假定为一个离散集合,并引入深度强化学习方法来处理复合储能设备的顺序决策问题。文献[40]提出了一种基于强化学习的自适应轻量级算法来获得储能的最优调度模型,有效降低系统运行成本和提高计算效率。文献[41]应用深度神经网络来学习历史电价信息,并优化微电网和公用电网之间的交换功率。现有的工作表明了深度强化学习方法可以用于解决储能调度问题,但是这些基于值的深度强化学习方法受到难以准确估计微电网储能设备充放电行动价值,从而导致微电网调度策略无法收敛到最优。因此,如何在微电网系统中准确估计储能设备的行为动作价值,实现微电网的最优经济运行,需要进一步研究。

随着多微电网系统的发展,深度强化学习也已广泛应用于多微电网协同能量调度领域。文献[42]将 Q 学习方法用于多微电网储能设备参与交易能源市场的投标策略。文献[43]将微电网的能量管理建模为马尔可夫决策过程,然后使用深度 Q 网络来降低设备的安全指标成本和提高可再生能源的利用率^[44]。文献[45]提出了一种双深度 Q 学习方法,以使正常运行状态下的利润最大化,并提高系统在

紧急情况下的可靠性。然而，深度 Q 网络的输出是离散的，这降低了微电网系统调度的准确性。为了解决上述深度强化学习方法输出的储能充放电动作离散的问题，文献[46-50]提出了一种基于深度确定性策略梯度算法的调度机制来最小化系统的能源成本。微电网经济的最优控制模型是通过软参与者—批评家算法训练得到^[51-54]。文献[55,56]提出了一种多智能体深度确定性策略梯度的多智能体强化学习方法，用于在隐私保护的情况下实现最优稳定性。然而，现有研究都没有考虑到多微电网协同调度过程中可能会遇到的部分信息缺失的情况。如何在非完全观测信息场景下实现多微电网能源调度优化问题仍然需要进一步研究。

随着电力系统的发展，综合能源系统是实现能源转型和构建可持续能源体系的关键途径，目前已有学者提出人工智能算法解决综合能源微电网多目标调度问题。文献[57]提出了一种综合能源系统混合时间尺度多目标强化学习低碳经济调度方法。文献[58]使用基于 Q 学习的多目标强化学习方法在智能家居中同时最小化电力成本和客户不满意度，文献[59]提出了一种基于多目标的能源管理策略的方法，设置了等效氢消耗、燃料电池耐用性退化和电池寿命退化的三维奖励函数，通过提出的多目标算法训练的调度，实现了对三个目标的优化。文献[60]提出了一种深度确定性策略梯度方法，以解决连续行动空间中考虑互补监管能力和经济运行的综合能源微电网多目标调度问题。上述算法主要是设置权向量，将多目标问题转化为单目标问题，无法找到调度优化的非凸帕累托（Pareto）前沿解集^[61]。因此，如何实现综合能源微电网多目标调度优化仍然是一个值得研究的问题。

1.3 主要研究内容

本文针对上述面临的调度优化问题，围绕基于深度强化学习算法的微电网调度优化，分别对微电网储能设备充放电调度、多微电网协同能源调度、综合能源微电网多目标调度三个方面进行研究。

1.3.1 微电网储能设备充放电调度

微电网中的储能设备作为系统安全稳定运行的保障，减少了可再生能源出力的波动性和间接性带给电力系统的不利影响。微电网储能设备充放电问题是一个序列决策的复杂问题。本文针对该问题，通过分析储能荷电状态和充放电之间的关系，建立储能设备充放电模型。考虑储能充放电寿命的使用次数建立储能设备

使用成本,基于实时电价、电池状态、可再生能源发电等因素分析储能充放电动作对系统的影响,建立微电网电池储能系统优化目标函数,考虑各设备特性以及系统功率平衡,进而对运行设备自身约束展开分析,最后通过提出的改进的深度Q网络算法对系统调度策略进行优化。

1.3.2 多微电网协同能源调度

微电网是将分布式发电装置和储能设备集成到电力系统中,可以提高系统能源利用率和运行经济性,但单一微电网系统在提高大规模可再生能源参透率上仍有一定局限性。多微电网系统通过自治管理和微电网间的能量互济,在提高电网运行的稳定性、经济性和能源利用率上优势明显,但多微电网系统中存在用户隐私安全或者微电网之间通信受限的情况,导致多微电网在非完全观测状态下难以实现最优运行的问题。本文针对该问题,考虑含有多储能设备和可再生能源发电装置的多微电网系统,分析多微电网之间协作关系,建立以系统总运行消耗电力最小化为目标的多微电网能源调度优化目标函数,通过提出的循环多智能体强化学习算法在部分可观测状态下对多微电网系统调度策略进行优化。

1.3.3 综合能源微电网多目标调度

综合能源微电网系统实现多元能源供应的充分互补,有效调动需求侧资源的响应潜力,可实现多能源系统间的横向多能互补与纵向协调发展,但是综合能源微电网系统模型复杂,难以优化多目标帕累托(Pareto)前沿的调度策略集。本文针对该问题,构建集成电、热、气等多种能源的综合能源微电网系统模型,根据室外室内温度、建筑材料、热水需求量、热水温度等因素建立热需求负荷模型,并根据电热负荷需求响应特征建立可转移、可消减的电热负荷模型,同时,考虑到需求响应会使用户用电习惯变化,从而导致用户产生不满意度,构建用户不满意度模型。此外,考虑系统运行排放的二氧化碳,建立系统环境污染模型。建立综合能源微电网多目标优化模型,构建多目标马尔可夫决策过程,通过所设计的多目标强化学习算法优化调度策略以得到 Pareto 前沿解集,最后应用模糊决策方法进行最优折衷解的选择,实现综合能源微电网多目标优化。

本文采取的技术路线如图 1-1 所示

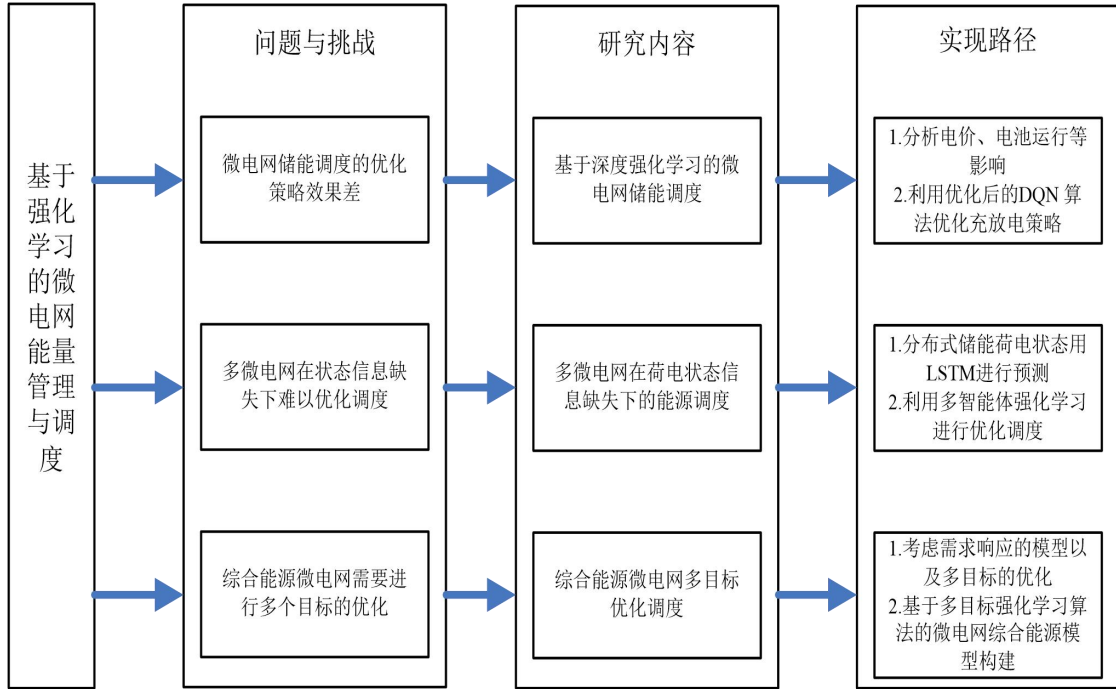


图 1-1 技术路线图

1.4 论文的组织结构

本文的内容分为六章，各章节的主要内容组织如下：

第一章，绪论。论述了课题的研究背景及意义。详细阐述了微电网调度优化研究现状，分析了当前研究存在的问题，并给出了本文的主要研究工作和组织安排。

第二章，强化学习相关理论及算法。介绍强化学习的基础理论、技术特点，并对深度强化学习的网络结构及应用方向进行了介绍和分析，为基于深度强化学习的微电网调度优化研究提供了基础。

第三章，基于深度强化学习的微电网储能调度优化。针对传统强化学习的微电网储能调度优化下评估行为价值偏差较大的问题，进行对神经网络的结构进行改进，提出基于精准价值评估的强化学习微电网储能调度方法。

第四章，基于深度强化学习多微电网协同能源调度优化。考虑到多微电网储能荷电状态缺失的场景，建立多微电网能源协同调度的部分可观测马尔可夫决策过程，提出基于循环神经网络强化学习的多微电网系统协同调度优化方法。

第五章，基于深度强化学习综合能源微电网多目标调度优化。建立考虑需求

响应的多目标综合能源调度模型,提出基于多目标强化学习的综合能源调度优化策略,实现对多个目标在对应目标权重下的最优调度。

第六章,总结与展望。总结本文的研究内容,并对研究中的不足之处进行讨论,展望未来的研究工作。

第 2 章 强化学习相关理论及算法

本章主要介绍强化学习的相关理论及算法。详细介绍了强化学习背景、马尔可夫决策过程、三种基础算法原理。为后文使用相关技术提供理论基础。

2.1 强化学习基础理论

2.1.1 强化学习背景

强化学习起源于行为心理学,最早提出强化概念的是美国心理学家爱德华·李·桑代克(Edward Lee Thorndike)的效应法则^[62],他指出动作可以从环境中获得反馈信号来促进自我强化,如果环境给出的信号很好,那么这个动作所选择的行为就会被强化,反之,这个动作所选择的行为就会被削弱。美国的马文·明斯基(Marvin Minsky)依据这一概念在 1954 年首次定义了强化学习^[63]。理查德·贝尔曼(Bellman)在研究最优控制问题时,提出了基于强化学习思想的动态规划方法,这一研究成果最终使学术界认识到强化学习理论中马尔可夫决策过程的重要作用^[64],并且强化学习在马尔可夫决策过程的支持下有了长足的发展。强化学习在之后的成长中开始融合其他学科领域的知识(例如统计学、控制理论等),逐步形成了一套完善的、在人工智能领域具有重要地位的理论体系,是当前国际上最有代表性的研究方向之一。强化学习的学习过程不同于监督学习和非监督学习,不需要依据带标签的数据,而是通过环境与主体交互之后反馈的强化信号对主体动作进行评价,然后通过评价促使动作达到预定目的,因此强化学习具有非常强的学习能力和对环境的适应能力。

2.1.2 马尔可夫决策过程

强化学习框架由智能体、环境、状态、动作、奖励组成,其本质是使智能体和环境交互,智能体基于观察到的环境状态选择动作,环境对该动作做出响应,智能体获得环境的反馈后调整下一步动作,最终实现智能体对环境的最优响应,如图 2-1 所示。

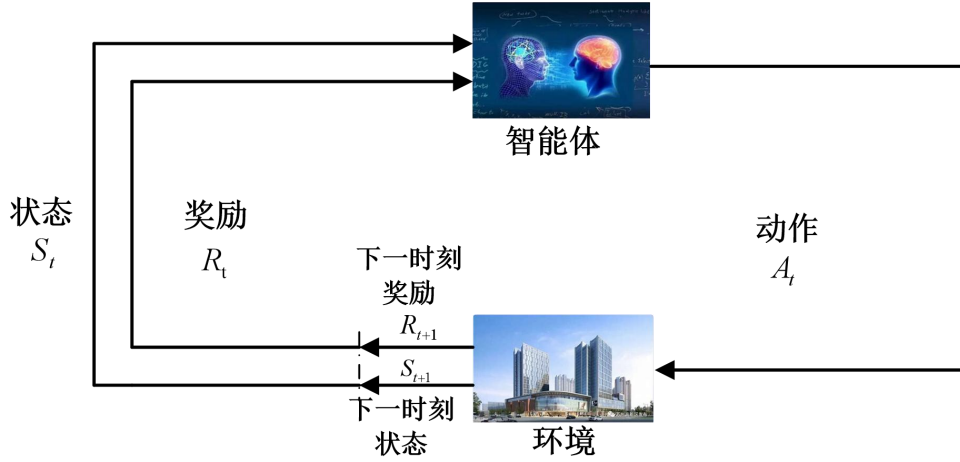


图 2-1 “智能体-环境”交互模式

在强化学习框架中，如果智能体的环境及其状态满足马尔可夫属性，则该强化学习任务称为马尔可夫决策过程（Markov decision process, MDP）^[65]。MDP 通常由 5 元组 (S, A, P_r, r, μ) 表示，其中： S 为环境状态空间，设 s_t 为 t 时段的环境状态； A 为动作空间，设 a_t 为 t 时段智能体的动作； P_r 为概率，设 $P_r(s_{t+1} | s_t, a_t)$ 为在状态 s_t 采取动作 a_t 后状态转移到 s_{t+1} 的概率； r 为即时奖励，设 r_t 为在状态 s_t 采取动作 a_t 后获得的即时奖励； μ 为策略，是状态到动作概率分布的映射，状态、动作和奖励序列构成策略的轨迹 τ 。假设未来奖励在每个时段的折扣因子为 γ ，在 T 时段终止时，累积奖励定义如式（2-1）所示。

$$R_t = \sum_{t=0}^{T-1} \gamma^t r_{t+1} \quad (2-1)$$

强化学习的目标是找到最优策略 μ^* ，使得所有状态的预期回报最大，如式（2-2）所示。

$$\mu^* = \operatorname{argmax} E_{\mu}[R_t | \mu] \quad (2-2)$$

式中： R_t 为 t 时段的累积奖励； argmax 表示求解最大化问题时的参数； $E_{\mu}[\cdot]$ 表示变量的期望值。

为了评估智能体所探索到的状态 $s \in S$ 的好坏程度，引入状态价值函数，用于表征从状态 s 开始，智能体采取策略 μ 的情况下对环境进行探索可以获得的累计奖励的期望值。状态价值函数可以表示为：

$$V_{\mu}(s) = E_{\mu}[R_t | s_t = s] \quad (2-3)$$

动作价值函数用于表征在状态 s 时，采取动作 a ，接下来智能体继续采取策略 μ 的情况下可以获得的累计奖励的期望值。动作状态价值函数可以表示为：

$$Q_{\mu}(s, a) = E_{\mu}[R_t | s_t = s, a_t = a] \quad (2-4)$$

利用强化学习寻找最优策略 μ^* 的问题可等价转化为寻找最大的动作价值函数问题，最大的动作价值函数 $Q^*(s, a)$ 对应的策略就是最优策略 μ^* ，如式（2-5）所示。

$$Q^*(s, a) = \max_{\mu} Q_{\mu}(s, a) \quad (2-5)$$

如果能够计算出最优状态价值函数或最优动作价值函数，则可以获得最优策略，求解最优策略需要对状态价值函数或动作价值函数进行准确的估计。

2.2 强化学习基础算法原理

2.2.1 动态规划算法

动态规划算法（Dynamic Programming, DP）可以分为策略迭代和价值迭代的方法，其中，策略迭代的方法分为两步：第一步进行策略评估，第二步进行策略提升。在策略评估阶段，根据价值函数的递推式可得：

$$\begin{aligned} V_{\mu}(s) &= E_{\mu}[r_t + \gamma V_{\mu}(s') | s_t = s] \\ &= \sum_{a \in A} \mu(a|s) (r_t + \gamma \sum_{s' \in S} P(s'|s, a) V_{\mu}(s')) \end{aligned} \quad (2-6)$$

使用上式进行迭代，可以得到当前策略的价值函数。此外，根据策略提升定理，可以证明如果对于任意 $s \in S$ ，满足 $Q^{\mu}(s, a) \geq V^{\mu}(s)$ ，则 $\mu'(s)$ 是优于 $\mu(s)$ 的，这里不再赘述。所以，可以使用贪心策略对策略进行更新：

$$\begin{aligned} \mu'(s) &= \operatorname{argmax}_a Q^{\mu}(s, a) \\ &= \operatorname{argmax}_a [r_t + \gamma \sum_{s' \in S} P(s'|s, a) V_{\mu}(s')] \end{aligned} \quad (2-7)$$

价值迭代的动态规划算法则是通过迭代逼近最优价值函数，在这个过程中策略也会向最优策略的方向更新：

$$v(s) = \max_{a \in A} \{r(s, a) + \gamma \sum_{s' \in S} P(s'|s, a) V_{\mu}(s')\} \quad (2-8)$$

2.2.2 蒙特卡洛算法

通过动态规划的迭代式可以看出，动态规划要求环境的转移概率已知，这在大多数场景下是不能被满足的。蒙特卡洛算法（Monte Carlo, MC），其实就是从等式（2-6）价值函数的定义出发，要想求得某个策略的价值函数，只需要依照该策略对环境进行多次采样，然后根据定义计算，可以用频率次数代替概率，这种方式所得到的结果近似于准确的估计。具体的：

$$Q(s, a) \leftarrow Q(s, a) + \frac{1}{N(s)} (G - Q(s, a)) \quad (2-9)$$

式中 G 表示累积回报奖励, $N(s)$ 为状态转化的次数。

2.2.3 时序差分算法

时序差分算法 (Temporal Difference, TD) 将动态规划算法和蒙特卡洛算法的思想进行了结合, 与蒙特卡洛算法不同的是, 时序差分算法不需要等到整个时期结束, 可以在智能体与环境交互的过程中估计动作价值函数, 将 (2-9) 式的 $1/N(s)$ 替换成固定步长 α , 则时序差分算法的更新公式为:

$$\begin{aligned} Q(s, a) &\leftarrow Q(s, a) + \alpha[G - Q(s, a)] \\ &= Q(s, a) + \alpha[r_t + \gamma Q(s', a') - Q(s, a)] \end{aligned} \quad (2-10)$$

上式中, $r_t + \gamma Q(s', a') - Q(s, a)$ 被称为时序差分误差 (Temporal Difference-error, TD-error)。

2.3 本章小结

本章主要介绍了强化学习的相关理论以及算法, 包括智能体和环境之间交互学习的马尔可夫决策过程以及动态规划算法、蒙特卡洛算法、时序差分算法三个基础强化学习算法, 为后续强化学习应用于微电网能量管理和调度优化领域的研究分析指明了研究方向。

第 3 章 基于深度强化学习的微电网储能充放电调度优化

3.1 引言

为实现二氧化碳净零排放目标，中国将大力发展可再生能源，增加相关设备装机容量。同时，发电的不确定性给配电网乃至整个电力系统运行期间的电压平衡带来了严峻的挑战。为了降低主电网波动性的不利影响，提出了由储能设备、新能源发电装置和柔性负荷组成的微电网。通过调度储能设备的充放电状态，可以平滑可再生能源发电出力的波动，同时也能降低微电网的运行成本。由于可再生能源发电量和负荷需求都是不确定的，如何优化储能设备的充放电策略成为微电网运行过程中的一个关键问题。目前已有大量深度强化学习算法应用于微电网系统能源调度中，但是神经网络预测微电网能源调度的动作价值的偏差会导致难以求解最优调度的问题。为了实现微电网系统能源调度优化策略，本章节将对微电网储能设备充放电调度优化建立模型，通过改进的深度 Q 网络强化学习算法对优化模型进行求解，最后通过仿真实验证明所提出的调度策略可以最小化系统运行成本。

3.1 系统模型描述与问题建模

3.1.1 系统模型描述

微电网的结构如图 3-1 所示。它由光伏发电设备、储能设备和负载组成。微电网的每个组成设备的细节描述如下。

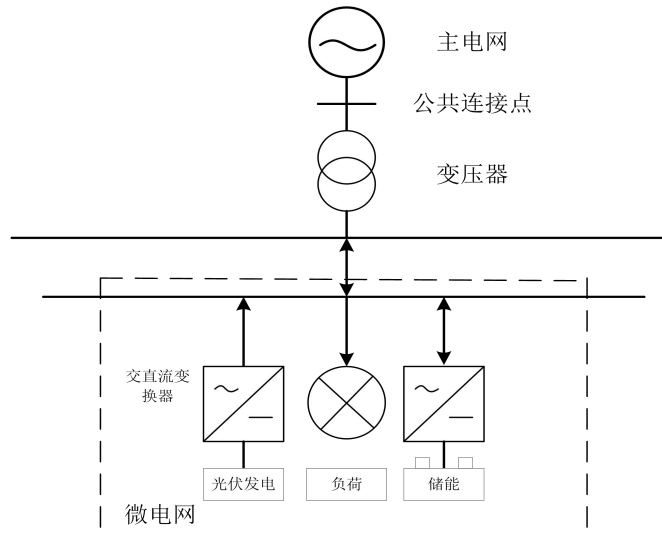


图 3-1 微电网结构图

1) 光伏发电设备

在本文中，我们考虑光伏发电为微电网系统的可再生能源。光伏发电量可由以下公式给出^[66]：

$$P_{PV}(t) = R(t)S\eta_{PV} \quad (3-1)$$

式中： $R(t)$ 为 t 时刻的太阳辐射强度， W/m^2 ； S 为布置的太阳能电池板的面积。太阳辐射强度与电池板面积的乘积为太阳能辐射功率，再乘以光伏发电的转换效率 η_{PV} ，得到 t 时刻光伏发电的功率。

2) 储能设备

储能设备由蓄电池组成，可与具有随机性和波动性的可再生能源协调运行，发挥削峰填谷的作用，保证微电网的可靠性和经济性。考虑蓄电池的充放电功率和储能荷电状态 SOC (State Of Charge) 的关系，储能设备充放电模型表达式为^[67]：

$$S_{t+1}^{soc} = S_t^{soc} + \frac{n_{ch}P_t^{ch}\Delta t}{E_{es}} - \frac{P_t^{dis}\Delta t}{n_{dis}E_{es}} \quad (3-2)$$

式中： S_t^{soc} 为 t 时段储能设备的 SOC； S_t^{ch} 和 S_t^{dis} 分别为 t 时段储能设备充电和放电状态值，其值为 0 表示关闭，为 1 表示开启； P_t^{ch} 和 P_t^{dis} 分别为 t 时段储能设备的充电和放电功率； n_{ch} 和 n_{dis} 分别为储能设备的充电和放电效率； E_{es} 为储能设备的额定容量； Δt 为时间间隔。

3) 储能设备退化成本

储能设备退化成本的主要考虑因素是设备的老化，这主要反映在电池单元可以充电和再放电的次数上。储能设备的退化成本模型设置为：

$$C(t) = \frac{C_E}{L_C} \quad (3-3)$$

式中： C_E 是电池的维护成本； L_C 是电池的生命周期。

3.1.2 问题建模

微电网储能系统充放电调度过程中，考虑到系统安全性和稳定性需要建立相应的约束条件。

1) 储能设备模型满足如下约束条件：

$$0 \leq P_t^{ch} \leq P_{max}^{ch}\mu \quad (3-4)$$

$$0 \leq P_t^{dis} \leq P_{max}^{dis}(1 - \mu) \quad (3-5)$$

$$S_{min}^{soc} \leq S_t^{soc} \leq S_{max}^{soc} \quad (3-6)$$

$$\mu \in \{0,1\} \quad (3-7)$$

其中 P_{max}^{ch} 和 P_{max}^{dis} 分别为储能设备的最大充电和放电功率，均为正值； S_{min}^{soc} 和 S_{max}^{soc} 分别为储能设备 SOC 的最小值和最大值。式（3-8）表示同一时段内储能设备不能同时进行充电和放电操作。

2) 微电网系统功率平衡约束条件：

$$P_t^{grid} + P_t^{pv} + P_t^{ES} = P_t^L \quad (3-8)$$

$$P_t^{ES} = P_t^{ch} - P_t^{dis} \quad (3-9)$$

其中 P_t^{grid} 是在时间 t 时与主电网的交换功率； P_t^{pv} 是时间 t 的 PV 有功功率； P_t^{ES} 是储能设备在时间 t 的输出功率； P_t^L 是时间 t 的负载需求。

微电网储能的运行调度以降低运行成本为目的，系统运行成本包括微电网的购电成本和储能设备维修成本组成，可以表述为：

$$\begin{aligned} obj &= \min[price(t)P_t^{grid} + C(t)] \\ s.t. &(3-4)-(3-9) \end{aligned} \quad (3-10)$$

其中 $price(t)$ 是外部电网在 t 时刻的电价。

3.2 微电网储能充放电调度优化算法设计

为了求解微电网调度优化，首先需要将优化问题转化为马尔可夫决策过程，然后利用强化学习算法求解。为了精准评估行为价值，本节使用最大均值差异深度 Q 网络（Maximum Mean Discrepancy-Deep Q learning Network, MMD-DQN）方法^[68]，该算法是在 DQN 算法的基础上对累积回报的分布进行建模，通过神经网络输出累积回报的分布信息，更加完整的得到累积回报的信息。

3.2.1 马尔可夫决策过程构建

通过设计相关的状态、动作、奖励函数将微电网储能系统调度优化问题转化为马尔可夫决策过程。具体的设计如下。

状态空间：状态包含负荷需求 P_t^L 、光伏输出功率 P_t^{pv} 、储能设备荷电状态 S_t^{soc} 和电价 $price(t)$ 。状态空间描述为：

$$S = [P_t^L, P_t^{pv}, S_t^{soc}, price(t)] \quad (3-11)$$

动作空间：储能设备调度动作指的是储能充放电的功率。动作空间描述为：

$$A = [P_t^{ES}] \quad (3-12)$$

奖励函数：奖励函数主要包括满足系统模型约束时微电网的运营成本，以及违反约束时的惩罚。奖励函数描述为：

$$r(t) = -(\text{price}(t)P_t^{\text{grid}} + C(t)) - \eta \quad (3-13)$$

其中 η 是超出约束的惩罚系数。

3.2.2 最大均值差异深度 Q 网络

由于状态转移存在随机性，智能体与环境之间也存在着随机性，经典强化学习方法的折扣累计回报 Z 是一个随机变量，给定策略 π 后，随机变量 Z 服从一个分布，称之为值分布。经典强化学习算法通过贝尔曼方程优化值分布的均值：

$$\tau^\pi Q(x, a) = E[R(x, a)] + \gamma E_{P, \pi} [Q(x', a')] \quad (3-14)$$

式中 τ^π 为通过策略 π 的贝尔曼操作符，是一个迭代更新运算； $E[\cdot]$ 为变量的期望值符号； $R(x, a)$ 为在状态 x 下执行动作 a 的奖励值； γ 为折扣因子； P 为状态转移概率； x' 和 a' 为下一时刻状态和动作。

值分布强化学习不是优化折扣累计回报 Z 的期望值，而是优化其分布，值分布贝尔曼操作符将一个随机变量映射成另外一个随机变量，即将一个分布映射成另外一个分布：

$$\tau^\pi Z(x, a) = R(x, a) + \gamma P^\pi Z(x', a') \quad (3-15)$$

式中 P^π 为下一个状态-动作处的概率分布。

对于一个给定的样本单元 (x, a, r, x', a') ，其中 x 是当前状态， a 是当前动作， r 是当前奖励， x' 是下一时刻状态， a' 是下一时刻动作。把随机变量 Z 的分布分成 N 个支点，每一个支点表示为 $(Z_\theta(x, a))_{i=1}^N$ ，每个支点（共 N 个）计算贝尔曼更新：

$$\tau Z_i = r + \gamma Z_\theta(x, a^*)_{i=1}^N, \forall i \in \{1, \dots, N\} \quad (3-16)$$

每个支点 $(Z_\theta(x, a))_{i=1}^N$ 通过公式(27)最小化最大均值差异（Maximum Mean Discrepancy, MMD）距离的均值来表示算法的损失函数：

$$L_{MMD^2}((Z_\theta(x, a))_{i=1}^N, (\tau Z_i)_{i=1}^N; k) \quad (3-17)$$

式中根据最小化经验 MMD 距离的均值来表示损失函数，用梯度下降的方法逼近最优值来实现 MMD 对值分布强化学习的求解。基于 MMD 方法的 DQN 强化学

习算法总流程为：

算法 3.1：最大均值差异深度 Q 网络算法

输入： 系统初始状态.

输出： 能源调度策略.

初始化：粒子数 N 、核函数 k （例如，高斯 RBF 核）、折损因子 $\gamma \in [0, 1]$ 、学习率 α 、重放缓冲区 M 、主网络 Z_θ 、目标网络 Z_{θ^-} 和策略 π （例如， ε -greedy 策略

$$w.r.t. Q^\theta(x, a) = \frac{1}{N} \sum_{i=1}^N Z_{\theta^-}(x, a)_i, \forall x, a).$$

1 重置 $\theta, \theta^- \leftarrow \theta$;

2 **For** $t=1, 2, \dots$ **do**

3 采取动作 $a_t \sim \pi(\cdot | x_t; \theta)$ ，获得奖励 $r_t \sim R(\cdot | x_t, a_t)$ 和状态 $x_{t+1} \sim P(\cdot | x_t, a_t)$;

4 存储回合数据 (x_t, a_t, r_t, x_{t+1}) 到重放缓冲区 M ;

5 从重放缓冲区 M 中随机抽取一批过渡样本 (x, a, r, x') ;

6 通过贝尔曼方程 $\tau Z_i = r + \gamma Z_{\theta^-}(x, a)_i, \forall i \in \{1, \dots, N\}$ 进行迭代更新，其中

$$a^* \leftarrow \operatorname{argmax}_{a' \in A} \frac{1}{N} \sum_{i=1}^N Z_{\theta^-}(x', a')_i;$$

7 更新主网络参数 $\theta \leftarrow \theta - \alpha \Delta_\theta L_{MMD^2}((Z_\theta(x, a)_i)_{i=1}^N, (\tau Z_i)_{i=1}^N; k)$;

8 定期更新目标网络 $\theta \leftarrow \theta^-$;

9 **End**

3.3 仿真实验与结果分析

3.3.1 仿真实验配置

在仿真实验中，以一天作为优化周期，每个时隙为 1 小时。深度神经网络的训练参数设置如下：每轮训练步长为 $T=24$ ，经验池容量为 10000。累积奖励函数分为 32 个概率区间。MMD-DQN 中神经网络的输入层有 4 个神经元，对应于 4 个环境状态；隐藏层有 64 的神经元；输出层有 32×11 个神经元，对应于离散动作空间中的 11 个动作。强化学习算法的超参数设置如下：神经网络的学习率 $\alpha=0.001$ ，折扣因子 $\gamma=0.95$ ，每批采样数为 64。本文已验证所设置训练参数可以使基准算法在该微电网调度模型中训练都能达到较好的效果。微电网储能系统主要参数如表 3.1 所示。

基准算法：深度 Q 网络(Deep Q-Network, DQN)^[69]，双重竞争深度 Q 网络

(Double Dueling Deep Q-Network, D3QN)^[70], 遗传算法(Genetic Algorithm, GA)^[71]

表 3.1 基本参数

参数名称	数值
储能设备的能源成本 C_E (元/周期)	171
储能设备的寿命周期 L_C (次/ $\text{kW}\cdot\text{h}$)	4996
储能容量上限 E_{es} ($\text{kW}\cdot\text{h}$)	5.60
储能充放电功率上限 P_t^{ES} ($\text{kW}\cdot\text{h}$)	1.12
充电效率系数 n_{ch}	0.95
放电效率系数 n_{dis}	1.02
储能设备的寿命周期率 η	0.99
储能设备的放电深度磨损成本效率 D_{od}	0.99

3.3.2 性能分析

图 3-1 比较了 MMD-DQN、DQN 和 D3QN 的训练过程。可以看出，三个算法的奖励和平均奖励在开始时迅速上升，在 500 个训练次数后逐渐收敛。结果表明，MMD-DQN 方法的收敛奖励值优于 DQN 和 D3QN 算法。

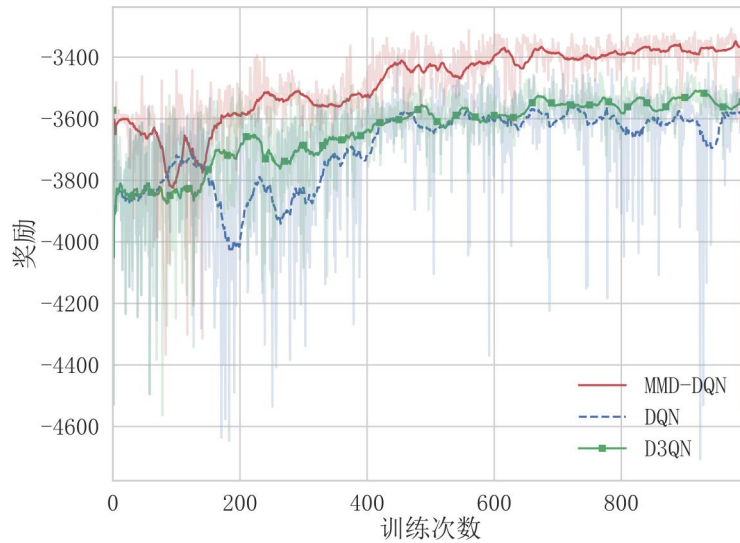


图 3-1 仿真算法训练过程

为进一步评估 MMD-DQN 算法的收敛性，定义如下指标：

$$\begin{cases} C_{MV} = \frac{\sum_{i=1}^N r^i}{N} \\ C_V = \frac{\sum_{i=1}^N (r^i - C_{MV})^2}{N} \\ C_{MAX} = \max(r^i) \\ C_{MIN} = \min(r^i) \end{cases} \quad (3-18)$$

式中： C_{MV} 为收敛均值； r^i 为完成第 i 次训练时的奖励，在实际中可对奖励进行

缩放； N 为训练总次数，每次训练的初始条件不同； C_V 为收敛方差； C_{MAX} 为收敛最大值； C_{MIN} 为收敛最小值。

表 3.2 收敛性评价指标对比

算法	C_{MV}	C_V	C_{MAX}	C_{MIN}
MMD-DQN	-3481.93	19574.54	-3307.28	-4367.52
D3QN	-3635.97	20963.74	-3427.19	-4563.73
DQN	-3717.62	41566.81	-3431.84	-4633.27

表 3.2 为三个算法训练 1000 次之后得到的性能指标，从表 3.2 中，我们可以看到，通过 MMD-DQN 算法获得的 C_{MV} 、 C_{MAX} 、 C_{MIN} 优于通过比较方法。这说明 MMD-DQN 算法具有较好的收敛性能。MMD-DQN 算法的 C_V 约为 DQN 算法的 47.09%，D3QN 算法的 93.37%。表明 MMD-DQN 算法的训练过程相对更稳定。

3.3.3 系统运行成本对比结果

为了验证基于 MMD-DQN 的微电网储能调度优化有效性，本节进行系统运行成本对比仿真。图 3-2 给出了一天内的光伏发电、负载、微电网购电价格的数据^[72]。

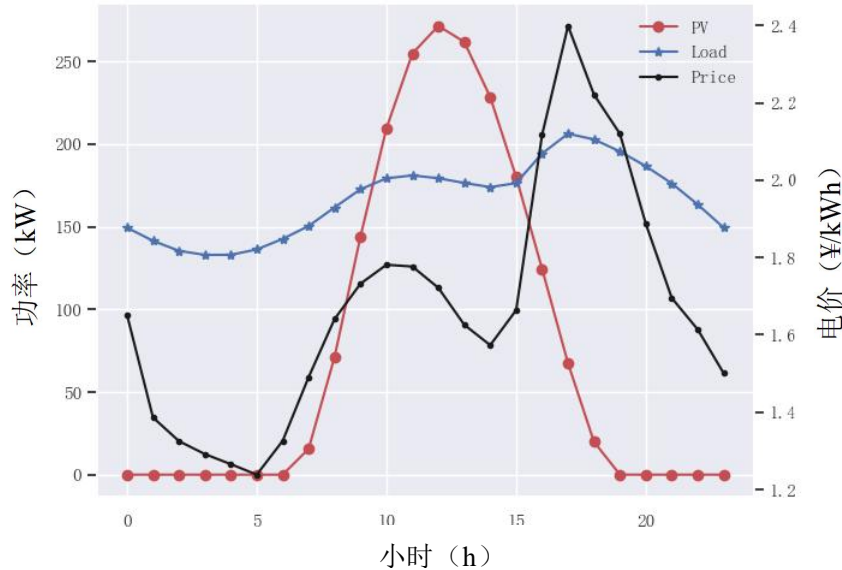


图 3-2 光伏发电、负载、微电网购电价格的仿真数据

图 3-3 示出了应用不同调度方法时储能设备的 SOC。结果表明，MMD-DQN 算法得到的充电/放电策略可以在价格低谷时增加充电量，在价格高峰期增加放电量。因此，微电网的运行成本大大降低。相反，D3QN 和 DQN 算法得到的策略在低电价期间没有尽可能多地充电，储能设备的容量没有得到充分利用。对于

GA 方法，它没有在高电价期间尽可能多地放电。因此，这三种比较方法的调度策略相对较差。

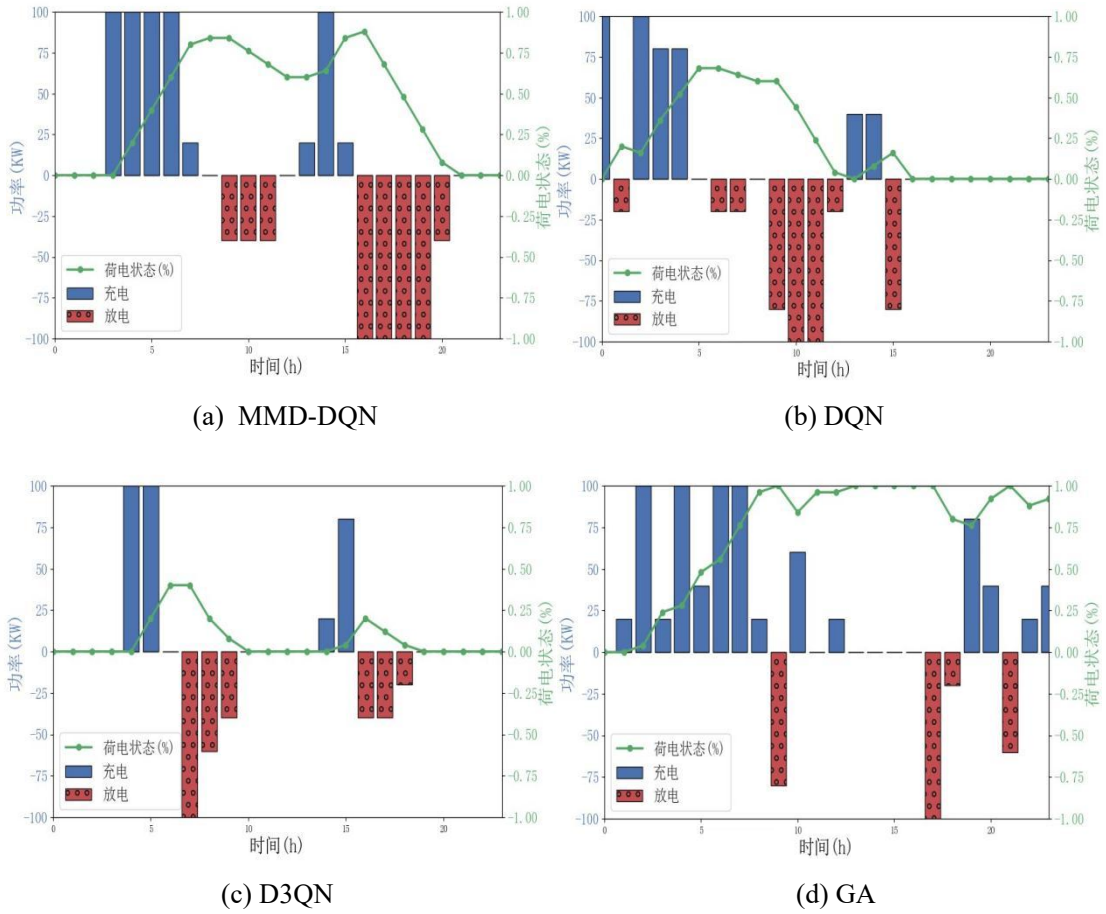


图 3-3 微电网储能充放电策略

表 3.3 显示了使用不同算法的一天运营成本。结果表明，通过基于 MMD-DQN 的策略获得的微电网总运营成本最低，分别比 D3QN、DQN、GA 获得的成本低 6.01%、9.29%、22.40%。

表 3.3 微电网运行成本

算法	储能维修成本/元	功率交换成本/元	总成本/元
MMD-DQN	146.64	3179.12	3325.77
D3QN	78.55	3460.12	3538.68
DQN	172.82	3493.42	3666.25
GA	128.31	4157.74	4286.05

上述实验结果表明，基于 MMD-DQN 的策略可以通过储能设备优化充放电功率有效降低微电网系统的运行成本。

3.4 本章小结

本章研究了微电网中的储能充放电功率调度问题,将调度问题被构建为马尔可夫决策过程。我们提出了一种基于 MMD-DQN 的值分布强化学习的微电网储能充放电调度优化模型来求解能量调度方案。仿真结果表明,与现有求解模型方法相比,该方法能够减少更多的系统运行成本。

第 4 章 基于深度强化学习的多微电网协同能源调度优化

4.1 引言

多微电网系统 (Multi-Microgrid System, MMG) 是一种先进的电力网络结构, 它由多个相互连接的微电网 (Microgrid) 组成。每个微电网本身是一个小型的电力系统, 能够独立运行或与更大的电网系统同步运行。多微电网系统的核心特点在于其模块化、自治管理和能量互济能力, 这些特点使其在提高电网的稳定性、经济性和能源利用率方面具有显著优势, 因此有大量学者对多微电网能源调度优化问题进行了研究。目前深度强化学习已广泛应用于多微电网协同能量调度领域, 但是很少考虑多微电网系统信息由于用户隐私问题或者通信受限导致部分缺失的情况。为此, 本章节构建多微电网协同能源调度优化模型, 通过提出的多智能体强化学习算法进行模型的求解, 最后进行仿真实验, 证明了所提出基于深度强化学习的多微电网协同能源调度策略能够减少系统运行的电能消耗。

4.1 系统模型描述与问题建模

4.1.1 系统模型描述

多微电网系统框架如图 4-1 所示。我们考虑三个微电网系统, 每个微电网都连接着对应的光伏 (PV) 板、风力涡轮机 (WT)、和负载需求单元。微电网可以在新能源发电供应的电能不足时从主电网中购买电功率, 在系统电功率过剩时可以把电功率储存在储能设备里。在多微电网系统在协同运行的情况下, 微电网之间也可以进行功率交换。储能可以通过充电和放电的能量调度来降低微电网系统来自主电网的功耗。建立了多微电网调度模式的目标函数, 在满足系统功率平衡的条件下, 使系统总功耗最小。

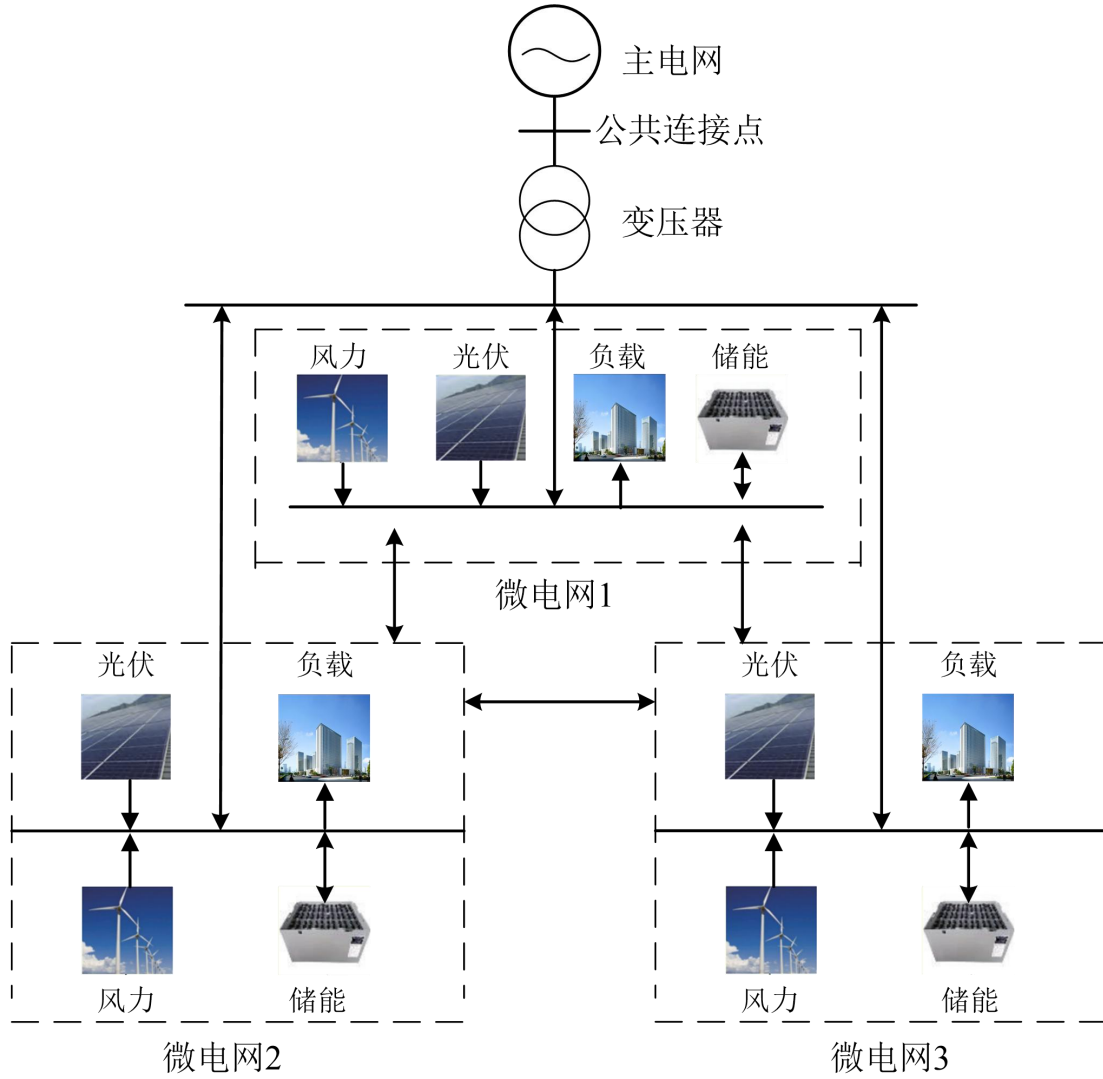


图 4-1 多微电网结构图

4.1.2 多微电网协同能源调度优化问题建模

主电网调度中心接受各个微电网信息，多微电网系统通过微电网中负载和新能源发电的特性进行互相协作，微电网电能短缺时不必先从主电网中获取电能，可以先从电能充盈的微电网获取电能来达到多微电网系统消耗主电网电能最小化的目标。多微电网系统目标函数表达式为：

$$f_M = \min \sum_{t=1}^T [\sum_{i=1}^N P_{i,t,grid} \Delta t] \quad (4-1)$$

$$P_{i,t,grid} = P_{i,t,ES} + P_{i,t,Load} - P_{i,t,RES} - \sum_{j \in N \setminus i} (P_{ij,t,input} - P_{ij,t,output}) \Delta t \quad (4-2)$$

$$f_{i,t,ES} = (P_{i,t,ch} + P_{i,t,dis}) \Delta t \quad (4-3)$$

$$f_{i,t,Load} = P_{i,t,L} \Delta t \quad (4-4)$$

$$f_{i,t,RES} = P_{i,t,PV} \Delta t + P_{i,t,WT} \Delta t \quad (4-5)$$

式中： $P_{i,t,grid}$ 为多微电网系统从主电网购买的电能； f_{ES} 、 f_{Load} 、 f_{RES} 分别为微电网 i 的在调度周期 T 的储能系统功率、负载功率、可再生能源功率； $P_{i,t,ch}$ 、 $P_{i,t,dis}$ 、 $P_{i,t,L}$ 、 $P_{i,t,PV}$ 、 $P_{i,t,WT}$ 分别为微电网 i 在时间 t 的储能系统充电功率、储能系统放电功率、负载消耗功率、光伏机组发电功率、风力发电功率； Δt 为步长； $P_{ij,t,input}$ 为微电网 i 在时间 t 时向微电网 j 的所获取的电能功率； $P_{ij,t,output}$ 为微电网 i 在时间 t 时向微电网 j 的所输出的电能功率。

其中可再生能源的功率约束主要考虑最大、最小功率约束：

$$P_{i,PV,min} \leq P_{i,PV}(t) \leq P_{i,PV,max} \quad (4-6)$$

$$P_{i,WT,min} \leq P_{i,WT}(t) \leq P_{i,WT,max} \quad (4-7)$$

式中 $P_{i,PV,min}$ 和 $P_{i,PV,max}$ 分别是微电网 i 中光伏发电机组最小和最大输出功率； $P_{i,WT,min}$ 和 $P_{i,WT,max}$ 分别是微电网 i 中风力涡轮机最小和最大输出功率。

储能约束主要考虑储能系统的充放电约束和自身容量约束^[67]：

$$0 \leq P_{i,t,ch} \leq P_{i,ch,max} \varepsilon_{i,t} \quad (4-8)$$

$$0 \leq P_{i,t,dis} \leq P_{i,dis,max}(1 - \varepsilon_{i,t}) \quad (4-9)$$

$$SOC_{i,t} = SOC_{i,t-1} + P_{i,t,ch} \eta_{i,ch} \Delta t / E - \frac{P_{i,t,dis} \Delta t}{\eta_{i,dis} E} \quad (4-10)$$

$$SOC_{i,min} \leq SOC_{i,t} \leq SOC_{i,max} \quad (4-11)$$

式中： $\varepsilon_{i,t}$ 取值为 0 或 1，其中 0 表示放电，为 1 表示充电； $P_{i,ch,max}$ 和 $P_{i,dis,max}$ 分别代表微电网 i 储能系统充电和放电的最大功率； $E_{i,t}$ 表示微电网 i 储能设备的额定容量； $\eta_{i,ch}$ 和 $\eta_{i,dis}$ 分别为充电和放电效率； $SOC_{i,t}$ 为微电网 i 在 t 时刻的储能系统荷电状态； $SOC_{i,min}$ 和 $SOC_{i,max}$ 分别代表微电网 i 储能系统荷电状态的上限和下限。

目标函数为微电网消耗主电网功率最小化，我们只考虑微电网向主电网获取电能的情况，其需要满足的约束为：

$$0 \leq P_{i,t,grid} \quad (4-12)$$

$$P_{i,t,RES} + P_{i,t,dis} - P_{i,t,ch} + P_{i,t,grid} = P_{i,t,L} \quad (4-13)$$

微电网之间以及与主电网的功率交互需要满足如下约束：

$$P_{i,t} = P_{i,t,grid} + \sum_{j \in N \setminus i} P_{ij,t,input} + \sum_{j \in N \setminus i} P_{ij,t,output} \quad (4-14)$$

$$0 \leq P_{ij,t,input} \leq f_{ij,t} \min\{P_{i,max}, P_{j,max}\}, i \neq j \quad (4-15)$$

$$0 \leq P_{ij,t,output} \leq (1 - f_{ij,t}) \min\{P_{i,max}, P_{j,max}\}, i \neq j \quad (4-16)$$

$$P_{ij,t,input} = P_{ji,t,output}, \quad i \neq j \quad (4-17)$$

式中 $P_{i,t}$ 为注入微电网 i 中的总功率； $f_{ij,t}$ 取值为0或1，1代表微电网 i 从微电网 j 中获取的功率，0代表微电网 i 向微电网 j 中输出的功率；式（4-17）表示微电网 i 从微电网 j 中获取的功率等于微电网 j 向微电网 i 中输出的功率，且 $i, j \in \{1, 2, \dots, N\}$ 。

4.2 多微电网协同能源调度优化算法设计

为了构建非完全观测下多微电网调度优化模型，首先需要将优化问题转化为部分可观测马尔可夫决策过程。然后使用长短期记忆神经网络预测缺失状态的信息，通过多智能体强化学习算法建立多微电网协同能源调度优化模型。

4.2.1 部分可观测马尔可夫决策过程

部分可观测马尔可夫决策过程(Partially Observable Markov Decision Process, POMDPs)问题可以看作是MDP问题的延伸,增加了观测集合和对应的观测模型。通常, POMDP问题可由多元组 $\langle S, A, P, R, O, \gamma \rangle$, 其中环境状态空间为 S ; $A = [a_1, a_2, \dots, a_N]$ 为所有环境智能体的联合动作空间, N 为环境中智能体个数; P 为联合状态转移函数; $R = [r_1, r_2, \dots, r_N]$ 为所有智能体联合奖励空间; $O = [o_1, o_2, \dots, o_N]$ 为所有智能体联合局部观测状态; $\gamma \in [0, 1]$ 为折扣因子, 表示对未来奖励程度的影响。其中联合状态转移函数为:

$$P(S, A, S'): S_t \times A_t \times S_{t+1} \rightarrow [0, 1] \quad (4-18)$$

式中 $[0, 1]$ 代表了在执行联合动作 A_t 的情况下, 由状态 $S_t \in S$ 转移到下一个状态 $S_{t+1} \in S'$ 的概率分布。智能体的奖励回报与状态转移相关, 且状态转移受联合动作影响, 则每个智能体的奖励回报可定义为:

$$r_i^\pi = E[r_{t+1} | S_t = s, a_i^t = a] \quad (4-19)$$

式中 $\pi = [\pi_1, \pi_2, \dots, \pi_N]$ 为联合策略, 每个智能体策略函数 π_i 表示为:

$$\pi_i(S, a): S_t \times a_i^t \rightarrow [0, 1] \quad (4-20)$$

进一步, 各智能体状态值函数与动作值函数为:

$$V_i^\pi(s) = E_i^\pi[r_{t+1} + \gamma V_i^\pi(S_{t+1}) | S_t = s] \quad (4-21)$$

$$Q_i^\pi(s, a) = E_i^\pi[r_{t+1} + \gamma Q_i^\pi(S_{t+1}, A_{t+1}) | S_t = s, A_t = a] \quad (4-22)$$

式中 $V_i^\pi(s)$ 和 $Q_i^\pi(s, a)$ 分别为状态值与动作值函数, 用于评价联合策略价值, 依据

贝尔曼方程进行迭代更新。

针对本节研究的多微电网协同能源调度模型，定义各个智能体设计状态空间、动作空间和奖励函数。

全局状态空间：全局状态空间是所有智能体状态空间的集合 S ，本文将当前所在的时刻 t 、负载消耗功率、可再生能源发电功率以及储能荷电状态作为状态，智能体 i 在 t 时刻获取的状态参数集合表示为 $s_{i,t} = \{t, P_{i,t,L}, P_{i,t,RES}, SOC_{i,t}\}$ ，考虑到所有智能体的全局观测，我们得到 $s_t = \{s_{1,t}, s_{2,t}, \dots, s_{n,t}\}$ 。

局部状态空间：由于微电网系统受到干扰或者带宽限制而导致无法获取全局状态空间信息，本文将智能体储能荷电状态作为受到干扰或者限制的信息，智能体 i 在 t 时刻获取的局部状态参数集合表示为 $o_{i,t} = \{t, P_{i,t,L}, P_{i,t,RES}\}$ ，考虑到所有智能体的局部观测，我们得到 $o_t = \{o_{1,t}, o_{2,t}, \dots, o_{n,t}\}$ 。

动作空间：动作空间是所有智能体的联合动作 A ，本文将微电网系统中储能设备的充放电功率作为动作，智能体 i 在 t 时刻动作表示为 $a_{ij,t} = \{P_{i,t,grid}, P_{ij,t,input}, P_{ij,t,output}\}$ ， $i \neq j$ ，则联合动作为 $a_t = \{a_{1,t}, a_{2,t}, \dots, a_{n,t}\}$ 。

奖励空间：奖励函数 $r_i(t)$ 分为两部分，一部分为满足约束时消耗主电网电量函数的相反数 $-F_i$ ，一部分为违反约束条件时产生的惩罚 $-\eta$ 。其中 η 为超出约束的惩罚系数，则奖励向量为 $r_t = \{r_{1,t}, r_{2,t}, \dots, r_{n,t}\}$ 。

4.2.2 多智能体强化学习算法

通信信息缺失不全问题是限制多智能体深度强化学习（Multiple Agent Deep Reinforcement Learning, MADRL）应用的主要瓶颈。传统 MADRL 方法虽然取得了效果，但其还存收敛速度慢等问题。在深度学习领域中，长短期记忆（Long Short Term Memory, LSTM）递归神经网络通过提取历史轨迹序列之间的特征，学习时间序列规律，并解决了网络训练出现的长期记忆消失问题。本文利用多智能体递归深度确定性策略梯度（Recurrent-Multi-Agent Deep Deterministic Policy Gradient, RMADDPG）的多智能体协同决策算法^[73]。

为了解决因为通信问题导致部分信息缺失的情况，提出了一种基于 LSTM 递归神经网络的多智能体演员-评论（actor-critic）模型，该模型包含生成策略的 actor 和评价策略的 critic 两个网络，两个网络在训练过程中不断相互交互和更新参数，比单独训练的网络更有效率。另外，LSTM 递归神经网络通过在训练步骤执行之前通过隐藏层 h 存储信息，扩展了 MADDPG 的 actor-critic 框架，这使得

多智能体在部分可观测状态下能够更有效地学习成为可能。本文所提出算法的训练框架如图 4-1 所示，伪代码如算法 4-1 所示。

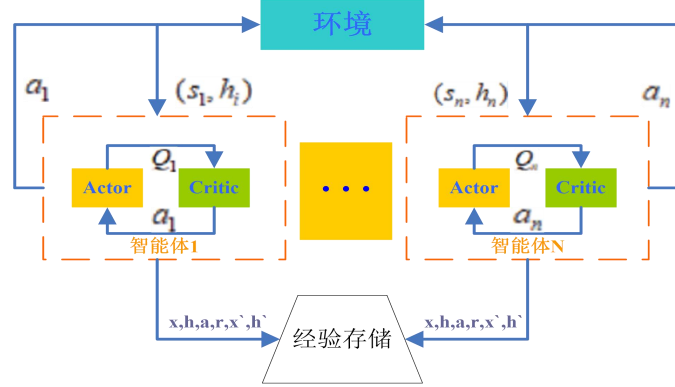


图 4-1 RMADDPG 算法框架图

算法 4-1：多智能体递归深度确定性策略梯度算法

输入： 系统初始状态.

输出： 能源调度策略.

初始化：参数 $\theta = \{\theta^1 \dots \theta^n\}$ 、 $w = \{w^1 \dots w^n\}$ 、actor 网络、critic 网络、经验池 R

- 1 **For** 训练次数=1 to M **do**
 - 2 **For** $t=1$ to T **do**
 - 3 初始化 h_0 ;
 - 4 **For** $i=1$ to n **do**
 - 5 根据公式 $a_t^i = \pi^i(h_{t-1}, o_t^i)$ 生成智能体 i 的动作;
 - 6 **End for**
 - 7 执行动作，并从环境中得到每个智能体的奖励 r_t^i 和新的观测 o_{t+1}^i ;
 - 8 更新历史信息编码 $h_t = \text{LSTM}(h_{t-1}, [O_{t-1}, A_{t-1}])$;
 - 9 存储经验 $h_0, o_1, a_1, r_1, h_1, o_2, a_2, r_2, \dots$ 到经验池 R;
 - 10 **End for**
 - 11 从经验池 R 中采样一个批次的样本;
 - 12 **For** $t=1$ to T **do**
 - 13 更新 Critic 网络:

$$LOSS = E_{h_{t-1}, o_t} [(Q^i(h_{t-1}, o_t, a) - y^i)^2]$$

$$y^i = r_t^i + \gamma Q(h_t, o_{t+1}, \mu^i(h_t, o_{t+1}^i));$$
 - 14 更新 Actor 网络
-

$$\nabla_{\theta^i} J(\theta^i) = E_{h_{t-1}, o_t} [\nabla_a Q(h_{t-1}, o_t, a)|_{a=\pi^i(h_{t-1}, o_t)} \nabla_{\theta^i} \pi^i(h_{t-1}, o_t^i)];$$

15 **End for**

16 **End for**

其中 i 表示第 i 个智能体; t 表示时刻; a_t^i 表示智能体 i 在 t 时刻的动作; π^i 表示智能体 i 采取的策略。在每一个训练周期内, 首先初始化历史信息编码 h_0 , 然后分别计算每个智能体需要执行的动作 a_t^i , 所有智能体根据自己的动作与环境互动, 收集每个智能体产生的及时奖励和下一时刻关于环境的观测, 之后将所有智能体的观测和动作传递给历史信息编码模块进行编码生成编码信息 h_1 , 并将智能体与环境交互产生的经验存储进经验池。在训练网络参数时, 首先从经验池中采集一个批次的经验, 然后依据经验序列分别计算 *Critic* 网络以及 *Actor* 网络的梯度, 并使用梯度下降法更新 *Critic* 网络参数, 依据策略梯度上升更新 *Actor* 网络的参数。

4.3 仿真实验与结果分析

4.3.1 仿真实验配置

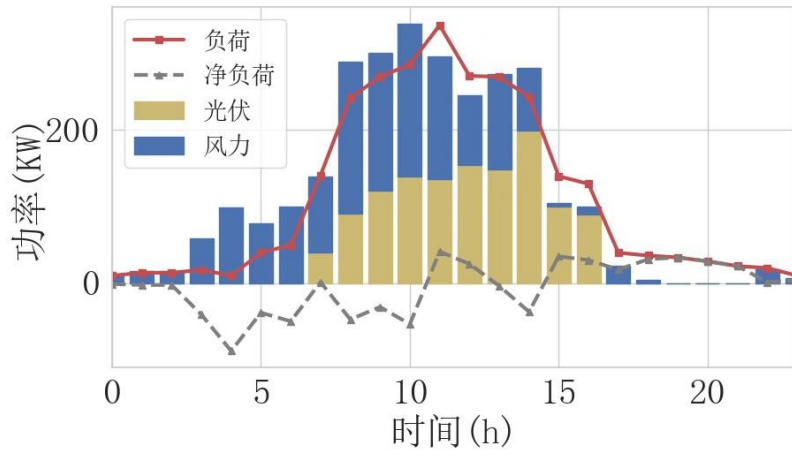
仿真系统中包括三个微电网, 即 MG1、MG2、MG3, 并且每个微电网包含 PV 面板、WT、负载和储能设备, 与外部电网相连接, 该仿真系统在储能设备荷电状态信息缺失的情况下运行。为了方便分析多微电网系统调度的优化, 我们将设置相同的微电网的模型参数。微电网模型的参数列于表 4.1。

表 4.1 基本参数

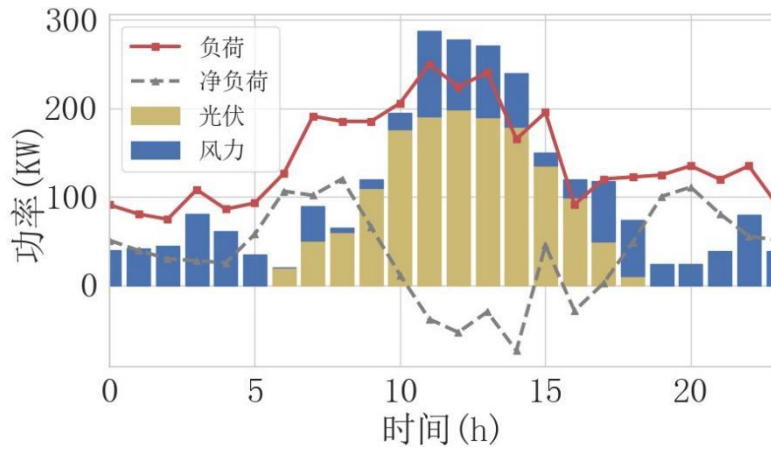
参数	值
$P_{ch,max}/P_{dis,max}$	50KW
$P_{ij,input}/P_{ij,output}$	25KW
E_{ES}	200KW
η_{ch}/η_{dis}	0.99/1.01
SOC_{min}/SOC_{max}	0.1/0.9

我们使用实际的光伏功率、风电功率和负荷需求数据作为实验数据^[74]。这些数据由传感器每小时收集一次。因此, 时间间隔被设置为一小时, 即 $\Delta t=1h$ 。上述参数分别表示在一个时间单元内的平均值。三个微电网在一天内的光伏功率和

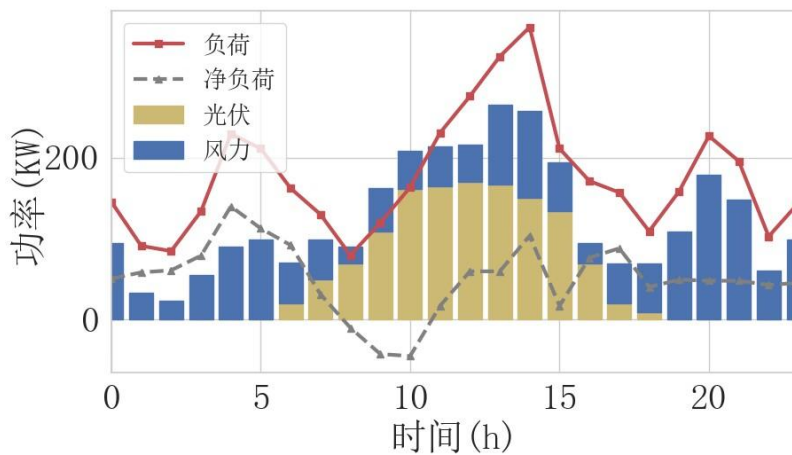
负载需求变化轨迹如图 4-2 所示，其中净负荷表示电力负载减去可再生能源输出功率的值。



(a) MG1



(b) MG2



(c) MG3

图 4-2 可再生能源和负载仿真数据

基准算法：多智能体深度确定性策略梯度(Multi-agent Deep Deterministic

Policy Gradient, MADDPG)^[75]、循环深度确定性策略梯度 (Recurrent Deep Deterministic Policy Gradient, DRPG)^[76]、独立深度 Q 网络 (independent Deep Q-Networks, IQL)^[77]。

4.3.2 算法性能对比

为了验证我们提出的基于算法的策略在 MMG 系统中的更好性能和协作操作系统模式的更好调度结果，我们将算法的性能与基准算法进行了比较，表 4.2 为算法在训练过程中所得到的性能指标。

表 4.2 算法收敛性能对比

算法	收敛值	平均值	最大值
RMADDPG	-47.28	-50.05	-46.23
MADDPG	-49.40	-51.21	-46.64
DRPG	-50.08	-53.78	-47.64
IQL	-54.31	-53.57	-50.69

从表 4.2 中观察到，在 POMDP 环境下，本文所提出的算法的收敛值分别比 MADDPG、DRPG、IRQL 算法高 4.29%、5.59%、12.94%，算法的平均和最大值同样优于其他算法。这表明在 POMDP 环境下，RMADDPG 算法比其他算法能学习到更好的策略。

4.3.3 仿真结果分析

通过对上述算法的性能评估和系统调度行为的分析，验证了在 POMDP 环境下，采用 RMADDPG 算法可以获取 MMG 系统的能量调度优化策略。在本小节中，我们将比较算法得到的能量调度策略时，微电网消耗的主电网电量的功耗。

图 4-3 表示各自算法得到的微电网消耗主电网电量的功耗。对于 MG1，采用 RMADDPG 算法时，微电网消耗的主网电量的功耗比 DRPG 和 IQL 算法低 20.14%和 69.41%。对于 MG2，通过上述分析，RMADDPG 算法的能量调度策略优于其他算法，功耗分别比 MADDPG、DRPG 和 IQL 算法低 4.46%、4.53%和 13.77%。对于 MG3，RMADDPG 算法的功耗也最低，分别比 MADDPG、DRPG 和 IQL 算法低 7.3%、2.2%和 6.5%。对于系统的总功耗，RMADDPG 的功耗分别比 MADDPG、DRPG 和 IQL 算法低 5.56%、4.3%和 12.95%。

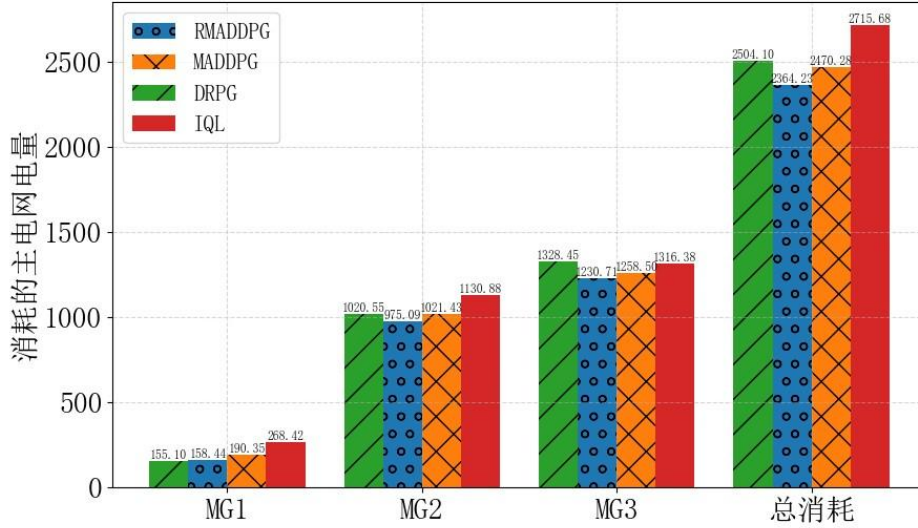


图 4-3 算法得到的功率消耗仿真结果

因此，基于 RMADDPG 的调度策略可以有效降低系统的运行功耗，实现 POMDP 环境下多微电网系统最小运行功耗的目的。

4.4 本章小结

在本章节中，我们研究了在缺失储能荷电状态信息下的多微电网协同能源调度优化问题，它被建模为部分可观测马尔可夫决策过程。我们使用长短期记忆神经网络预测缺失信息，构建基于 RMADDPG 强化学习的多微电网调度优化策略。仿真结果表明，与现有调度优化策略相比，该策略能够减少更多的系统用电量。

第 5 章 基于深度强化学习的综合能源微电网能源调度优化

5.1 引言

随着能量管理技术的不断发展,多种供能方式能够集成为综合能源微电网系统,实现多种能源的协调联合互补。由于综合能源的电网集成了电、热、气等多种能源的优点,因此它能显著地提高自身的供能能力。除此以外,综合能源微电网系统通过对多能源进行联合运行调度,能够实现对具有高不确定性与波动性的风电、光伏等新能源发电的大规模消纳。此外,通过采用需求响应策略,使用柔性负载来缓解调峰压力并提高综合能源微电网系统的可靠性。考虑到改变用户用电习惯会导致客户舒适度的降低,本章节将构建系统运行成本、用户不满意度、污染气体排放三个目标的模型,并且提出多目标强化学习算法对该模型进行求解,最后通过仿真证明该策略实现多目标优化的有效性。

5.1 系统模型描述与问题建模

5.1.1 系统模型描述

综合能源微电网系统是由多种能源耦合形成的为微电网系统,图 5-1 为综合能源微电网的框架。综合能源微电网设备包括风力涡轮机(WT)、光伏(PV)、微型涡轮机(MT)、热交换器(HE)、吸收式制冷机(AC)、电锅炉(EB)、电制冷机(EC)、燃气锅炉(GB)、储能设备和负载。冷热电三联供(CCHP)包括微型涡轮机、热交换器、吸收式制冷机。系统中的储能设备包含三种类型,即电池存储(BS)、热存储(TS)和冷存储(CS)。综合能源微电网负载包括电、热、冷负载。综合能源微电网可以连接到主电网购买电力,也可以连接到天然气公司购买天然气以产生加热/冷却能量。

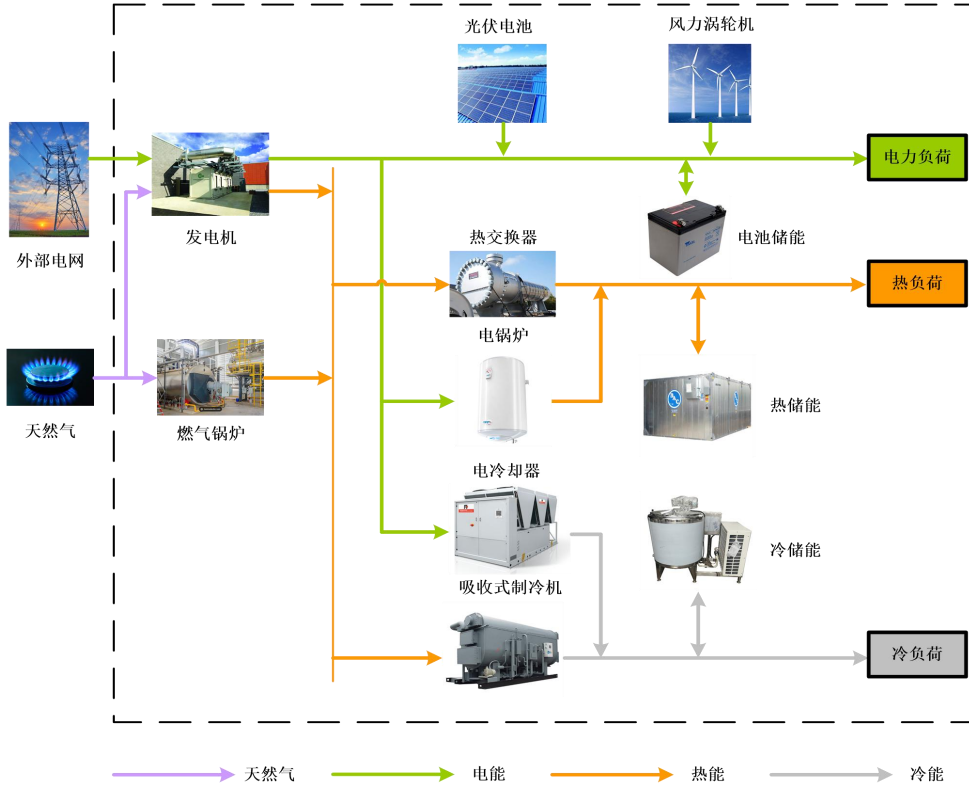


图 5-1 综合能源微电网系统结构

1) 冷热电三联供模型:

冷热电三联供由三部分组成：微型涡轮机、热交换器和吸收式制冷机。微型涡轮机可以提供电力和热能。热交换器/吸收式制冷机可以将微型涡轮机产生的废热转化为热能/冷能。因此，冷热电三联供可以显著提高能源效率。微型涡轮机的输出电功率和产生的热/冷功率之间的关系如下所示^[78]：

$$Q_{MT}^t = \frac{P_{MT}^t(1-\eta_{MT}^t-\eta_L)}{\eta_{MT}^t} \quad (5-1)$$

式中 Q_{MT}^t 是微型涡轮机在时隙 t 中产生的热/冷功率； P_{MT}^t 是微型涡轮机产生的电力； η_{MT}^t 是微型涡轮机功率输出的效率； η_L 是热损失比。

综合能源微电网系统中热交换器/吸收式制冷机的热/冷输出可以表示为：

$$\begin{cases} Q_{MTH}^t = \eta_{rec}\eta_{HE}Q_{MT}^t \\ Q_{MTC}^t = \eta_{rec}\eta_{AC}Q_{MT}^t \end{cases} \quad (5-2)$$

其中 Q_{MTH}^t 和 Q_{MTC}^t 是热交换器和吸收式制冷机产生的热和冷功率； η_{HE} 和 η_{AC} 是热交换器和吸收式制冷机的加热和冷却系数； η_{rec} 是废热的回收率；微型涡轮机的天然气消耗量显示如下：

$$V_{MT}^t = \frac{P_{MT}^t}{\eta_{MT}^t L_{NG}} \Delta t \quad (5-3)$$

式中 L_{NG} 是低热值； Δt 是调度的时间间隔。

2) 电锅炉和电冷却器模型

电锅炉/电冷却器可以将电能转换成热/冷电能，并用于配合能量转换，以提高能量供应的灵活性。电锅炉/电冷却器的能量转换效率在每个时隙 t 中被设置为常数，其描述如下：

$$\begin{cases} Q_{EB}^t = \eta_{EB} P_{EB}^t \\ C_{EC}^t = \eta_{EC} P_{EC}^t \end{cases} \quad (5-4)$$

式中 Q_{EB}^t 和 C_{EC}^t 是电锅炉和电冷却器的热输出功率和冷输出功率； η_{EB} 和 η_{EC} 是电锅炉和电冷却器的冷热系数； P_{EB}^t 和 P_{EC}^t 分别代表电锅炉和电冷却器的电功率消耗。

3) 燃气锅炉

燃气锅炉可以通过消耗天然气来产生热/冷能。燃气锅炉的输出功率可以表述为：

$$\begin{cases} Q_{GBH} = \eta_{GB} V_{GBH} L_{NG} \\ Q_{GBC} = \eta_{GB} V_{GBC} L_{NG} \end{cases} \quad (5-5)$$

式中 Q_{GBH} 和 Q_{GBC} 是燃气锅炉提供的热能和冷能； V_{GBH} 和 V_{GBC} 分别是热交换器和吸收式制冷机消耗的天然气量； η_{GB} 表示燃气锅炉的能量效率。

4) 电力负荷需求响应

电力负载包括固定负载（BL）和柔性负载。固定负载代表了大部分不能改变的电力负载。柔性载荷包括三部分：可消减载荷（CL）、可移动载荷（SL）和可转移载荷（TL）。

可削减的负荷（例如，照明、空调）可以根据客户的需求来减少。每次切割功率的限制为：

$$\alpha_{cut}^t P_{cut}^{min} \leq P_{cut}^t \leq \alpha_{cut}^t P_{cut}^{max}, t_{cut}^{min} < t < t_{cut}^{max} \quad (5-6)$$

式中 α_{cut}^t 是二进制变量， $\alpha_{cut}^t=1$ 表示具有可缩减功率， $\alpha_{cut}^t=0$ 表示在时间 t 时没有可缩减负载； P_{cut}^t 是负载削减功率； P_{cut}^{min} 和 P_{cut}^{max} 分别表示可缩短周期的最小和最大限制。

可削减的负载将对电源造成不便，并对客户产生补偿成本。客户在时间 t 的补偿成本建模如下：

$$\begin{cases} C_{cut}^t = \chi_{cut} \alpha_{cut}^t P_{cut}^t, & t_{cut}^{min} \leq t \leq t_{cut}^{max} \\ C_{cut}^t = 0, & t < t_{cut}^{min}, t > t_{cut}^{max} \end{cases} \quad (5-7)$$

式中 χ_{cut} 是负荷削减的单位补偿成本。

可移动负载（例如干衣机、洗衣机）可以从高峰时间移动到低谷时间。允许的负载转移次数限制如下：

$$\sum_{t=1}^T \alpha_{shift}^t \leq N_{shift}^{max} \quad (5-8)$$

式中 α_{shift}^t 是二进制变量， $\alpha_{shift}^t=1$ 表示具有可移位负载， $\alpha_{shift}^t=0$ 表示没有可移位负载。 N_{shift}^{max} 是允许的最大负载转移次数。时间 t 的可移动负载的总补偿成本可以表示为：

$$\begin{cases} C_{shift}^t = \chi_{shift} \alpha_{shift}^t P_{shift}^t, & t_{shift}^{min} \leq t \leq t_{shift}^{max} \\ C_{shift}^t = 0 & , t < t_{shift}^{min}, t > t_{shift}^{max} \end{cases} \quad (5-9)$$

式中 $[t_{shift}^{min}, t_{shift}^{max}]$ 表示允许负载转移的时间间隔； χ_{shift} 表示负载转移的单位补偿成本； P_{shift}^t 表示时间 t 的移动功率。

可转移载荷可以在不改变可转移载荷总和的情况下，在一定时间段内分配载荷。因此，可转移可以实现调峰和填谷。每个时间的电力约束可以表示如下。

$$\alpha_{trans}^t P_{trans}^{min} \leq P_{trans}^t \leq \alpha_{trans}^t P_{trans}^{max}, \quad t_{trans}^{min} < t < t_{trans}^{max} \quad (5-10)$$

式中 α_{trans}^t 是二进制变量， $\alpha_{trans}^t=1$ 表示开始传输负载，否则不是传输负载； P_{trans}^{min} 和 P_{trans}^{max} 分别是负载转移的最小和最大功率； $[t_{trans}^{min}, t_{trans}^{max}]$ 是负载转移的允许周期的周期间隔； P_{trans}^t 表示时间 t 的传输功率。

时间 t 的可转移负载的总补偿成本可以表示如下：

$$\begin{cases} C_{trans}^t = \chi_{trans} \alpha_{trans}^t P_{trans}^t, & t_{trans}^{min} \leq t \leq t_{trans}^{max} \\ C_{trans}^t = 0 & , t < t_{trans}^{min}, t > t_{trans}^{max} \end{cases} \quad (5-11)$$

式中 χ_{trans} 表示负载转移后的单元补偿成本；

5) 冷热负载

在本文中，我们将达到热水和室内空气的预期温度所消耗的功率视为热/冷负荷需求，并可以建立模型，其中总热负荷需求包含热水和室内空气需求的温度，总冷负荷需求仅考虑室内空气需求的温度。

热水可以通过综合能源微电网中的储水箱储存，但不考虑管道中抽水流量动态和环境温度的影响。我们认为储水箱中的水总是满的，当消耗一些热水时，可以装满相应的冷水。热水的温度主要取决于三个因素：加热功率、水的热容和注入冷水的体积。热水的模型可以表示如下^[79]：

$$\theta_w^{t+1} = \frac{\rho_w V_{cold}^t (\theta_{cw} - \theta_w^t) C_w + \rho_w V_w \theta_w^t + Q_{wat}^t \Delta t}{\rho_w V_w C_w} \quad (5-12)$$

式中 θ_w^{t+1} 和 θ_w^t 是在 $t+1$ 和 t 时储水箱中热水的温度； V_{cold}^t 是在时间 t 注入冷水的体积； θ_{cw} 是注入水箱的冷水的温度； Q_{wat}^t 是通过综合能源微电网提供给水箱的热能。 ρ_w 、 V_w 和 C_w 分别是水的密度、储水箱的容量和水的比热容。

空气传热与建筑材料有关，可以模拟如下：

$$Q_B^t = \frac{1}{R} [Q_{out}^t - Q_{in}^t] \quad (5-13)$$

式中 R 是建筑材料的耐热性； Q_{out}^t 和 Q_{in}^t 是时间 t 的室外空气和室内空气的温度； Q_B^t 在综合能源微电网中能量传递为热能时为负，在传递冷能时为正。

当室内空气被转移时，需要热能来提供适当的热能 Q_{air}^t 以将温度保持在舒适的范围内。室内空气温度的动力学建立如下：

$$C_{air} \frac{d\theta_{in}^t}{dt} = Q_B^t + Q_{air}^t \quad (5-14)$$

式中 C_{air} 是空气比热。结合等式（5-13）和（5-14），由热能提供的室内温度可以重新描述如下：

$$\theta_{in}^{t+1} = \theta_{in}^t e^{-\frac{1}{RC_{air}}} + (RQ_{air}^t + \theta_{out}^t)(1 - e^{-\frac{1}{RC_{air}}}) \quad (5-15)$$

类似地，当传递冷功率 C_L^t 时，室内空气温度的动态可以描述如下：

$$\theta_{in}^{t+1} = \theta_{in}^t e^{-\frac{1}{RC_{air}}} + (-RC_L^t + \theta_{out}^t)(1 - e^{-\frac{1}{RC_{air}}}) \quad (5-16)$$

为了避免同时供应热能和冷能，所需功率的约束如下：

$$Q_{air}^t C_L^t = 0 \quad (5-17)$$

5.1.2 问题建模

本文考虑了综合能源微电网的多目标优化问题，其目的是最小化系统的运行成本、二氧化碳排放、客户不满意度。

1) 目标函数

第一个目标是降低综合能源微电网的运营成本。总运营成本包括购买电力和天然气的成本、设备的维护成本、设备的启动成本以及分配灵活负载的成本。目标函数可以表述如下。

$$\min F_1 = \sum_{t=1}^T (F_B^t + F_O^t + F_S^t + F_D^t) \Delta t \quad (5-18)$$

$$F_B^t = [k_{grid} P_{grid}^t + k_{gas} (V_{MT}^t + V_{GB}^t)] \Delta t \quad (5-19)$$

$$F_O^t = \sum_{i=1}^I K_i P_i^t \Delta t \quad (5-20)$$

$$F_S^t = \sum_{j=1}^M \max\{0, U_{DG,j}^t - U_{DG,j}^{t-1}\} C_{ST,j} \Delta t \quad (5-21)$$

$$F_D^t = (C_{cut}^t + C_{shift}^t + C_{trans}^t) \Delta t \quad (5-22)$$

$$\begin{cases} P_{BS}^t = P_{BS,ch}^t + P_{BS,dis}^t \\ Q_{TS}^t = Q_{TS,ch}^t + Q_{TS,dis}^t \\ Q_{CS}^t = Q_{CS,ch}^t + Q_{CS,dis}^t \end{cases} \quad (5-23)$$

式中 F_B^t 是在时隙 t 中购买电力和天然气的成本； F_O^t 是设备维护成本； F_S^t 是设备启动—停止成本； F_D^t 为灵活负荷调度成本； P_{BS}^t 、 Q_{TS}^t 和 Q_{CS}^t 分别是电池储能、热储能和冷储能的充电/放电功率； P_{grid}^t 是在时隙 t 中从主电网到综合能源微电网的电力； k_{grid} 是系统从主电网购买电能的价格； k_{gas} 是天然气的价格； k_i 是设备的维护成本 i ； $U_{DG,j}^t$ 是设备 j 的状态， $U_{DG,j}^t=1$ 表示设备 j 打开，而 0 表示设备关闭； $C_{ST,j}$ 是设备 j 的起止成本；

灵活负荷的调度会扰乱客户正常的用电习惯，给客户带来不便，产生对用电的不满意度。客户不满意度目标函数建模如下：

$$\min F_2 = \sum_{t=1}^T (\lambda_1 f_E^t + \lambda_2 f_T^t) \quad (5-24)$$

$$f_E^t = (\mu_1 \frac{P_{cut}^t}{P_{cut}^{max}} + \mu_2 \frac{P_{shift}^t}{P_{shift}^{max}} + \mu_3 \frac{P_{trans}^t}{P_{trans}^{max}}) \quad (5-25)$$

$$f_T^t = (\sigma_1 \frac{|\Delta\theta_w^t|}{(\theta_w^{max} - \theta_w^{min})} + \sigma_2 \frac{|\Delta\theta_{in}^t|}{(\theta_{in}^{max} - \theta_{in}^{min})}) \quad (5-26)$$

式中 f_E^t 是时间 t 的柔性电力负荷的不满意函数； f_T^t 是在时间 t 使用加热/冷却能量的不满意函数； λ_1 和 λ_2 是一阶权重系数； μ_1 、 μ_2 和 μ_3 是柔性电力负荷的次要权重系数； σ_1 、 σ_2 是使用加热/冷却能量的次要权重系数； θ_w^{max} 和 θ_w^{min} 是热水温度的上限和下限； θ_{in}^{max} 、 θ_{in}^{min} 是室内温度的最大和最小极限； $\Delta\theta_w^t$ 、 $\Delta\theta_{in}^t$ 分别是室内热水和空气在实际和舒适水平之间的温度偏差。

本文主要考虑二氧化碳作为污染排放气体，二氧化碳排放的目标函数可表述如下：

$$\min F_3 = \sum_{t=1}^T (\xi_{MT} P_{MT}^t + \xi_{GB} Q_{GB}^t + \xi_{grid} P_{grid}^t) \quad (5-27)$$

式中 ξ_{MT} 、 ξ_{GB} 和 ξ_{grid} 分别是微型涡轮机、燃气锅炉和外部电网的二氧化碳排放系数。

2) 模型约束

在系统运行过程中，必须考虑具体的约束条件，如系统的功率平衡、每组分布式发电机的功率限制以及储能设备的充放电限制。

综合能源微电网系统产生的电力和从主电网获得的电力必须满足电力需求。同样，冷热电三联供和能量转换装置产生的热/冷电力应满足热/冷负荷。在不考

虑网络损耗的情况下，时隙 t 中发电和负荷之间的电力平衡可以表示如下：

$$\begin{cases} P_{WT}^t + P_{PV}^t + P_{grid}^t + P_{GT}^t = P_{EL}^t + P_{EB}^t + P_{EC}^t + P_{BS}^t \\ P_{EL}^t = P_{base}^t + P_{cut}^t + P_{shift}^t + P_{trans}^t \end{cases} \quad (5-28)$$

式中 P_{WT}^t 和 P_{PV}^t 分别是风力涡轮机和光伏发电的输出功率； P_{EL}^t 表示时隙 t 中的总电力负荷。

热负荷的平衡约束为

$$\begin{cases} Q_{MTH}^t + Q_{GBH}^t + Q_{EB}^t = Q_L^t + Q_{TS}^t \\ Q_L^t = Q_{wat}^t + Q_{air}^t \end{cases} \quad (5-29)$$

式中 Q_L^t 是时间 t 的总热负荷。

冷负荷的平衡约束为

$$C_{EC}^t + Q_{MTC}^t + Q_{GBC}^t = C_L^t + C_{CS}^t \quad (5-30)$$

式中 C_L^t 是时间 t 的总冷负荷。

设备提供的功率应保持在上限和下限内，以防止损坏。发电机设备和能量转换操作对风力涡轮机功率、光伏功率、冷热电三联供功率、燃气锅炉热能和电锅炉/电制冷机热能有最大和最小限制。

$$\begin{cases} P_i^{min} \leq p_i^{min} \leq p_i^{max} \\ Q_i^{min} \leq q_i^{min} \leq q_i^{max} \end{cases} \quad (5-31)$$

5.2 综合能源微电网多目标调度优化算法设计

为了构建微电网综合能源储能多目标调度策略，首先需要将优化问题转化为多目标马尔可夫决策过程。然后使用多目标强化学习算法求解多目标综合能源调度优化帕累托（Pareto）前沿解集，最后通过模糊决策选取合适的折衷解，实现多个目标对应目标权重下的最调度优化策略。

5.2.1 多目标马尔可夫决策过程

多目标马尔可夫决策过程（Multiple Objective Markov Decision Process, MOMDP）可以定义为元组 $\langle S, A, p, r, \mu, \gamma \rangle$ 其中 S 是状态空间； A 是动作空间； $p(\cdot|s, a)$ 是从当前状态 $s \in S$ 执行动作 $a \in A$ 后到下一个状态的转移概率； $r = [r_1, r_2, \dots, r_m]$ 是具有 m 个不同目标的多目标奖励函数向量； μ 是初始状态分布概率； $\gamma \in (0, 1)$ 是折扣因子。 s_t 、 a_t 和 r_t 是对应于时间 t 的状态、动作和向量奖励的随机变量。集合 π 表示所有可能的策略，其中 $\pi \in \Pi$ 是将状态映射到动作分布

的函数。对于给定的状态—动作 (s, a) ，策略的多目标动作值函数可以定义如下：

$$Q^\pi(s, a) = E_\pi[\sum_{i=0}^{\infty} \gamma^i r_{t+i} | S_t = s, A_t = a] \quad (5-32)$$

式中 $Q^\pi(s, a)$ 是 m 维的向量函数； E_π 是在 π 的轨迹下的期望。 v_π 是随着策略 π 的交互轨迹上的多目标值向量函数：

$$v^\pi = E_{s_0 \sim \mu}[Q^\pi(s_0, \pi(s_0))] \quad (5-33)$$

式中 $v^\pi(s, a)$ 是 m 维的向量函数， v_j^π 是 π 上第 j 个目标上的值。

通常，在多目标优化中，目标经常相互冲突，很难同时获得最优解。多目标规划的解集由一组非支配多目标值函数组成的 Pareto 边界表示。帕累托边界是：

$$F = \{v^\pi | \exists \pi' \text{ s.t. } v^\pi \succ_p v^{\pi'}\} \quad (5-34)$$

其中 $\succ_p$ 是 Pareto 支配关系。多目标优化问题的解是一组策略 π ，这使得 v^π 处于帕累托前沿。

效用函数可以通过将多目标值向量 v^π 映射到标量来评估策略 π 的性能。效用函数定义为 $u(v^\pi, w) = v^\pi$ ，它使用一组权重 w 线性组合每个目标下的策略值。权重 w 的每个元素是每个目标的相对重要性，所有权重元素之和等于 1。对于给定的常数 w ，MOMDP 奖励函数可以定义为 $r_w(s, a, s') = r(s, a, s') * w$ 。Pareto 前沿的有限凸子集可以定义为给定线性效用函数 u 的凸覆盖集 (Convex Covering Set, CCS)。CCS 确保集合中存在一个最优的策略，并且对应于任何线性偏好 w 。CCS 的表达式如下：

$$CCS = \{v^\pi \in F | \exists w \text{ s.t. } \forall v^{\pi'} \in F, v^\pi w \geq v^{\pi'} w\} \quad (5-35)$$

对于综合能源微电网能量调度问题，MOMDP 中状态、动作和奖励的空间设计表示如下：

状态空间：综合能源微电网系统状态包括风力涡轮机和光伏发电的输出功率、储能系统的 SOC、负荷需求、室内温度、热水量、主电网电价和天然气价格。时间 t 处的状态空间如下：

$$s_t = [P_{WT}^t, P_{PV}^t, SOC_{BS}^t, SOC_{TS}^t, SOC_{CS}^t, P_{base}^t, P_{cut}^t, P_{shift}^t, P_{trans}^t, Q_{air}^t, C_L^t, V_{cold}^t, k_{grid}^t, k_{gas}^t] \quad (5-36)$$

动作空间：综合能源微电网能量调度动作包括可调度分布式发电的加热/冷却功率、转换能量单元的输出功率、储能设备的充放电功率、燃气锅炉产生的加热/冷却功率和灵活负载功率。时隙 t 中的动作空间表示如下：

$$a_t = [Q_{MTH}^t, Q_{MTC}^t, P_{EB}^t, P_{EC}^t, P_{BS}^t, P_{TS}^t, P_{CS}^t, Q_{GBH}^t, Q_{GBC}^t, P_{cut}^t, P_{shift}^t, P_{trans}^t] \quad (5-37)$$

综合能源微电网多目标能源调度优化包括同时最小化总运营成本、二氧化碳排放、客户不满意度。因此，奖励函数是一个三维向量函数，每个元素对应一个目标。将多目标优化问题转化为求解最大累计奖励函数。其中时刻 t 的奖励函数如下：

$$\begin{cases} r_t = [r_1^t, r_2^t, r_3^t] \\ r_1^t = -(F_B^t + F_O^t + F_S^t + F_D^t) \\ r_2^t = -(\lambda_1 f_E^t + \lambda_2 f_T^t) \\ r_3^t = -(\xi_{MT} P_{MT}^t + \xi_{GB} P_{GB}^t + \xi_{grid} P_{grid}^t) \end{cases} \quad (5-38)$$

式中 r_1^t 、 r_2^t 和 r_3^t 分别对应于在 t 时刻总运营成本、客户不满意度、二氧化碳排放的奖励函数。

5.2.2 多目标强化学习算法

在本节中，我们提出了多目标双延迟深度确定性策略梯度 (Multiple Objective Twin Delayed Deep Deterministic Policy Gradient, MOTD3) 算法，它可以在任何权重下训练最优策略。所提出的 MOTD3 算法通过在学习过程中获得最大累积回报来学习最优策略。图 5-2 描述了基于 MOTD3 的能量管理策略的框架。

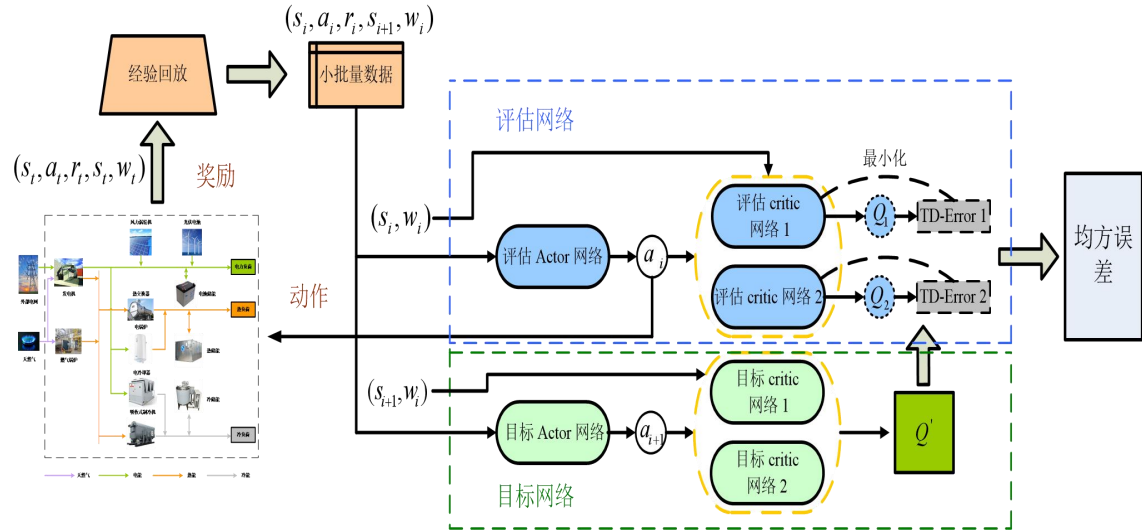


图 5-2 MOTD3 算法结构

所提出的 MOTD3 算法扩展了单目标 DRL 算法，通过深度神经网络 (Deep Neural Network, DNN) 输出动作值函数。所提出的算法结合了两类 DNN：

评估网络和目标网络。评估网络参数表示为 θ ，而目标网络参数表示为 θ' 。评估网络和目标网络都由一个行动者网络和两个批评者网络组成。DNN 的输入包括状态 s 、动作 a 和偏好权重 w 。DNN 的动作值函数输出可以定义如下：

$$Q_{\theta}^{\pi}(s, a, w) = E_{\pi}[\sum_{i=0}^{\infty} \gamma^i r_{i+1} w | s_t = s, a_t = a, \theta] \quad (5-39)$$

动作值函数由贝尔曼方程更新：

$$y(r, s', w) = rw + \gamma \min_{i=1,2} Q_{\theta'_i}(s', a', w) \quad (5-40)$$

式中 s' 和 a' 分别表示下一个状态和下一个动作。 θ'_1 和 θ'_2 是两个评论家网络的参数。评论家网络选择最小行动值函数，以减轻价值 Q 的高估。通过最小化损失函数来更新网络参数：

$$L(\theta_i) = (Q_{\theta_i}(s, a, w) - y(r, s', w))^2 \quad (5-41)$$

参与者网络输出动作 a 。我们提出的算法为动作增加了噪声：

$$a = \pi_{\varphi}(s', w) + \varepsilon, \quad \varepsilon \sim \text{clip}(\mathfrak{N}(0, \sigma), -c, c) \quad (5-42)$$

式中 ε 是范围为 $(-c, c)$ 的动作噪声，它执行的动作保持在一个合理的范围内； $\mathfrak{N}(0, \sigma)$ 是均值为 0，方差为 1 的高斯分布。参与者网络参数的策略梯度更新如下：

$$\nabla_{\varphi} J(\varphi) = \sum \nabla_a Q_{\theta_i}(s, a, w)|_{a=\pi_{\varphi}(s, w)} \nabla_{\varphi} \pi_{\varphi}(s, w) \quad (5-43)$$

式中 φ 是评估 actor 网络参数。MOTD3 方法使用软更新方法来更新参数，其表示如下：

$$\begin{cases} \theta'_i \leftarrow \tau \theta_i + (1 - \tau) \theta'_i \\ \varphi'_i \leftarrow \tau \varphi_i + (1 - \tau) \varphi'_i \end{cases} \quad (5-44)$$

式中 τ 是更新系数； φ 是目标参与者网络的参数。MOTD3 算法的实现过程流程图如下。

算法 5-1：多目标双延迟确定性策略梯度算法

输入： 系统初始状态.

输出： 能源调度策略.

初始化 actor 网络和 critic 网络参数 θ_1 和 θ_2 ，初始化经验池 D ，权重样本分布 D_w

1 重置目标网络 $\theta'_1 \leftarrow \theta_1$, $\theta'_2 \leftarrow \theta_2$;

2 **For** 训练次数=1 to M **do**

3 初始化 s_0 ;

4 线性权重样本 $w \sim D_w$;

5 **For** $t=1$ to T **do**

```

6      选择探索噪音的动作:
           $a = \pi_{\varphi'}(s, w) + \varepsilon, \varepsilon \sim \text{clip}(\mathfrak{N}(0, \sigma), -c, c);$ 
7      观测奖励  $r$  和下一状态  $s'$ ;
8      存储元组  $(s, a, r, s')$ ;
9      从 D 中抽取小样本元组;
10      $a' = \pi_{\varphi'}(s', w) + \varepsilon, \varepsilon \sim \text{clip}(\mathfrak{N}(0, \sigma), -c, c);$ 
11      $y(r, s', w) = rw + \gamma \min_{i=1,2} Q_{\theta'_i}(s', a', w);$ 
12     更新 critics:
           $\theta_i \leftarrow \text{argmin}_{\theta_i} N^{-1} \sum (y - Q_{\theta_i}(s, a, w))^2;$ 
13     If t mod d then:
14         更新参数  $\varphi$ :
               $\nabla_{\varphi} J(\varphi) = N^{-1} \sum \nabla_a Q_{\theta_i}(s, a, w)|_{a=\pi_{\varphi}(s, w)} \nabla_{\varphi} \pi_{\varphi}(s, w);$ 
15         更新目标网络:
               $\theta'_i \leftarrow \tau \theta_i + (1 - \tau) \theta'_i;$ 
               $\varphi'_i \leftarrow \tau \varphi_i + (1 - \tau) \varphi'_i;$ 
16     End for
17 End for

```

5.3 仿真实验与结果分析

5.3.1 仿真参数

实验数据主要包括风电、光伏的输出功率、各类电力负荷的需求、热水的消耗量（注入的冷水量）、室外温度、主电网的电力购买价格、燃气公司的天然气购买价格以及综合能源微电网系统的参数^[80]。热需求主要通过热水满足，冷需求主要考虑夏季冷却室内空气温度；考虑热水的热需求和室内空气温度，冬季不需要室内空气温度的冷需求；考虑到天气变化和客户舒适度，夏季和冬季的预期室内空气温度和热水具有不同的特征。

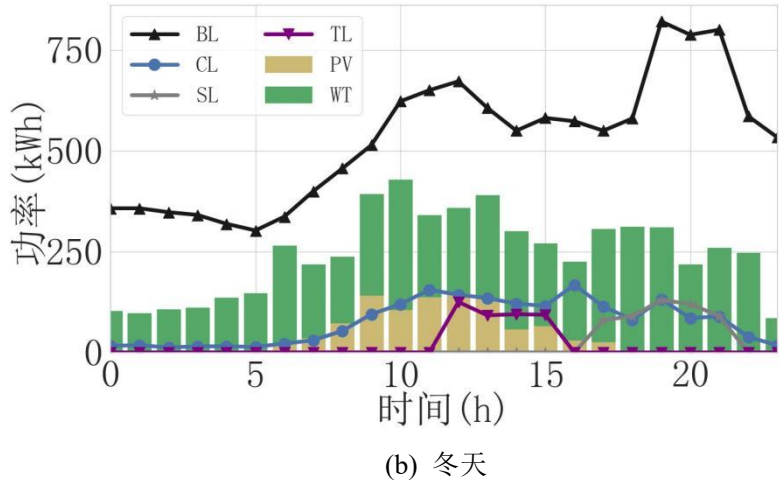
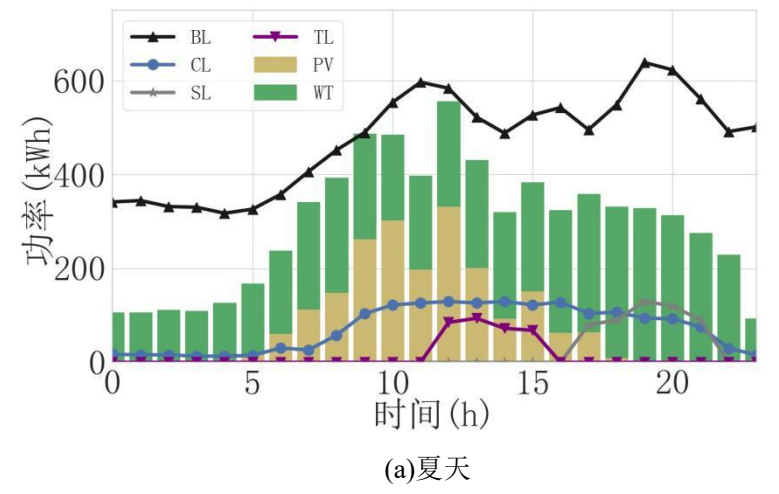
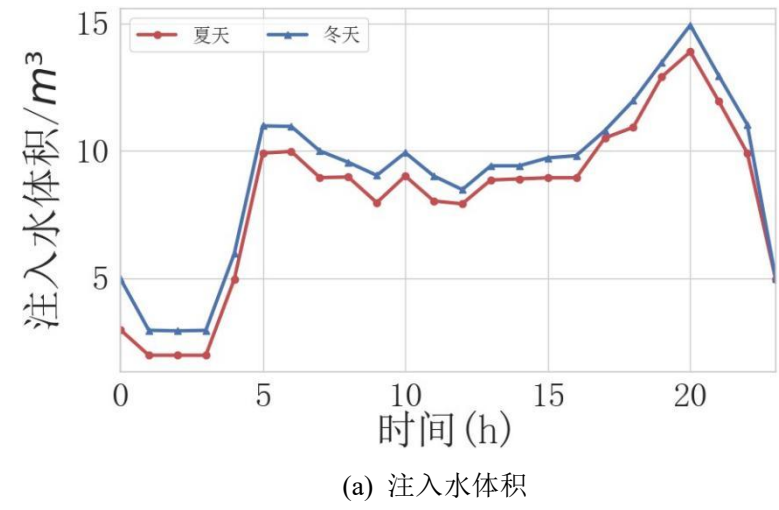
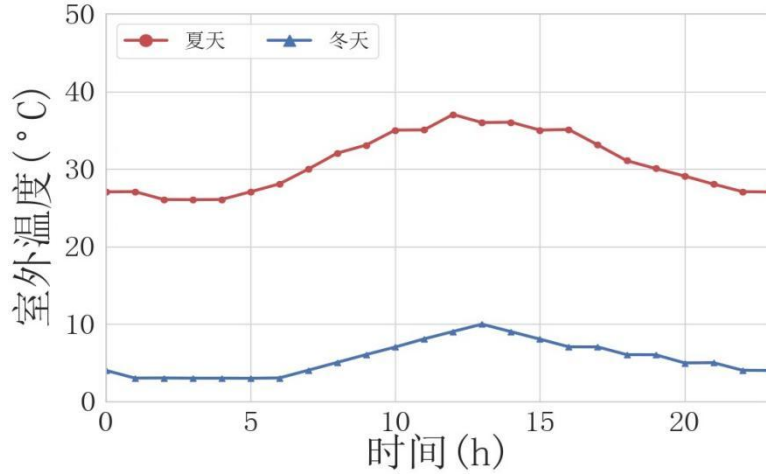


图 5-3 仿真数据





(b) 室外温度

图 5-4 温度和热水需求

WT、PV 和电力负荷需求的预测数据如图 5-2 所示。注入冷水的体积和室外空气的温度如图 5-3 所示。一阶权重系数 λ_1 和 λ_2 被设置为 0.5。次要权重系数 μ_1 、 μ_2 和 μ_3 被设置为 0.333，并且 σ_1 和 σ_2 为 0.5。使用时间的电价数据见表 5.1。天然气的单位购买价格为 $3/m^3$ ，低热值为 $9.7 \text{ kW}\cdot\text{h}/m^3$ 。综合能源微电网系统的参数和二氧化碳排放量见表 5.2。综合能源微电网系统中元件的功率限制显示在表 5.3 中，综合能源微电网中储能设备的参数显示在表 5.4 中。表 5.5 显示了热水和室内空气温度的参数。在本文中，我们设定微型涡轮机、热交换器、吸收式制冷机的起止成本分别为 1.94、1.32 和 1.36。

表 5.1 电价

类型	时间	价格
谷价	24: 00-07: 00	0.25
平价	07: 00-10: 00, 15: 00-18: 00, 21: 00-24: 00	0.53
峰价	10: 00-15: 00, 18: 00-21: 00	0.58

表 5.2 设备参数

参数	值	参数	值
k_{MT}	0.05(\$/kWh)	ξ_{MT}	184(g/kWh)
k_{GB}	0.03(\$/kWh)	ξ_{GB}	232(g/kWh)
k_{EC}	0.09(\$/kWh)	ξ_{grid}	981(g/kWh)
k_{EB}	0.09(\$/kWh)	η_{EB}	0.98
k_{BS}	0.026(\$/kWh)	η_{EC}	3
k_{TS}	0.013(\$/kWh)	η_{HE}	1.28
k_{CS}	0.013(\$/kWh)	η_{AC}	1.36
k_{PV}	0.035(\$/kWh)	η_{GB}	0.85
k_{WT}	0.035(\$/kWh)	η_{MT}	0.45
η_{nec}	0.65	η_L	0.20

表 5.3 设备功率限制

变量	范围	变量	范围
P_{grid}^t	[0,800]	Q_{GBH}^t	[0,3000]
P_{MT}^t	[0,500]	Q_{GBC}^t	[0,500]
P_{WT}^t	[0,350]	Q_{MTH}^t	[0,300]
P_{PV}^t	[0,400]	Q_{MTC}^t	[0,200]

表 5.4 储能参数

类型	BS	TS/CS	类型	BS	TS/CS
τ	0.001	0.002	λ_{min}	0.2	0.2
η_{ch}/η_{dis}	0.95/1.05	0.98/1.02	λ_{max}	0.9	0.9
$P_{ch,max}$	150	200	SOC_0	0.2	0.2
$P_{dis,max}$	150	200	E_{ES}	500	800

表 5.5 热水和室内温度参数

变量	单位	值	变量	单位	值
R	$^{\circ}\text{C}/\text{kW}$	0.18	$\theta_{w,satis}^{sum}$	$^{\circ}\text{C}$	60
ρ_w	kW/m^3	10e3	$\theta_{w,satis}^{win}$	$^{\circ}\text{C}$	70
v_w	m^3	90	θ_{cw}	$^{\circ}\text{C}$	10
C_w	$\text{kWh}/\text{kg} \cdot ^{\circ}\text{C}$	1.116e-3	$\theta_{in,min}^{sum}$	$^{\circ}\text{C}$	21
C_{air}	$\text{kWh} \cdot ^{\circ}\text{C}$	0.525	$\theta_{in,max}^{sum}$	$^{\circ}\text{C}$	26
$\theta_{w,min}^{sum}$	$^{\circ}\text{C}$	50	$\theta_{in,min}^{win}$	$^{\circ}\text{C}$	25
$\theta_{w,max}^{sum}$	$^{\circ}\text{C}$	70	$\theta_{in,min}^{win}$	$^{\circ}\text{C}$	30
$\theta_{w,min}^{win}$	$^{\circ}\text{C}$	60	$\theta_{in,satis}^{sum}$	$^{\circ}\text{C}$	24
$\theta_{w,max}^{win}$	$^{\circ}\text{C}$	80	$\theta_{in,satis}^{win}$	$^{\circ}\text{C}$	28

为了评估我们提出的 MOTD3 算法的有效性,我们将其与其他三种算法进行了比较: 1) 多目标粒子群优化算法 (MOPSO) [81]; 2) 非支配排序遗传算法 II (NSGA-II) [82]; 以及 3) 包络多目标 Q 学习算法 (Envelop) [83]。

5.3.2 性能对比

通过不同的评价指标来评价我们的方法应用于多目标能量调度问题的有效性。收敛性、多样性和一致性是多目标优化问题最基本的标准。本文引入超体积来比较算法的性能。

超体积度量可以评估通过多目标规划获得的 Pareto 前沿集的收敛和分集性能,如图 5-3 所示。超体积度量根据 Pareto 边界的支配部分相对于选定参考点 R 的体积来计算算法的性能。具体来说,我们将 x 作为 Pareto 边界的非支配解的集合,其中 $x_i \in X$ 。参考点 R 和解 x 中的 x_i 构造超立方体 v_i 的对角。超体积由体积较大的所有超立方体组成,其值越大表示收敛性和多样性越好。

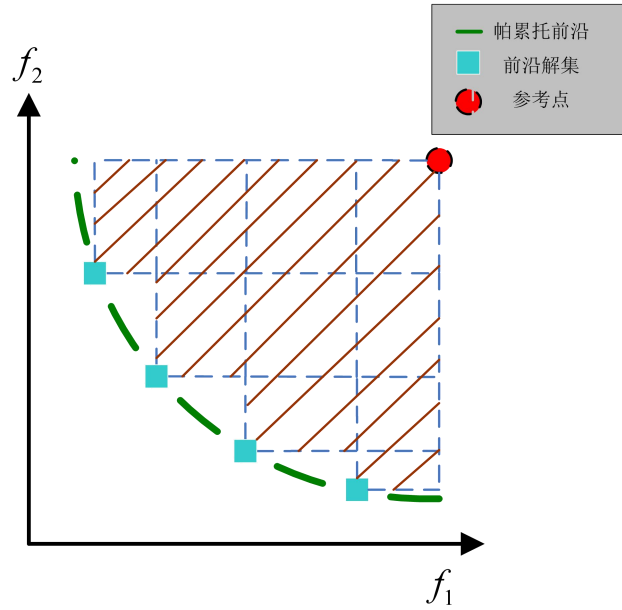


图 5-3 帕累托图

表 5.6 显示了超容量的算法性能模拟结果，其中我们选择参考点 $R=[15e3, 15, 22e3]$ 。MOTD3 算法具有最高的超体积值，这意味着它比其他算法具有更好的收敛性和多样性。由于综合能源微电网能量调度的动作空间很大，Envelop 离散空间的算法使得难以有效地学习多目标优化策略。

表 5.6 算法性能对比结果

季节	算法	超体积值(e6)
夏天	MOTD3	3102.33
	MOPSO	2059.88
	NSGA- II	111.81
	Envelop	2433.13
	MOTD3	2035.81
冬天	MOPSO	1159.14
	NSGA- II	948.23
	Envelop	1204.28

5.3.3 仿真结果

对于多目标优化模型，为了平衡三个目标，我们使用模糊决策方法从 Pareto 前沿找到最终的折衷解。根据最大隶属度对应的权重系数，从 Pareto 边界中选择最终的折衷方案。

在图 5-4 中，根据一天的热水和室内空气温度比较了各种算法。MOTD3 算

法在 05:00 之前保持热水的高温，并逐渐将其降低到最低要求极限。MOTD3 确保室内空气温度稳定在舒适的水平。由于热水的热负荷高于室内空气的热负荷，MOTD3 的调度策略通过减少热水供应来降低运行成本，并通过满足室内空气的热负荷来提高客户舒适度。相反，MOPSO 和 NSGA-II 算法具有相对较大的温度波动，这不能保证运行成本最小化和客户满意度。包络算法始终保持热水温度稳定，减少了客户不满意度，但由于水和室内空气的热负荷较高，增加了运营成本。

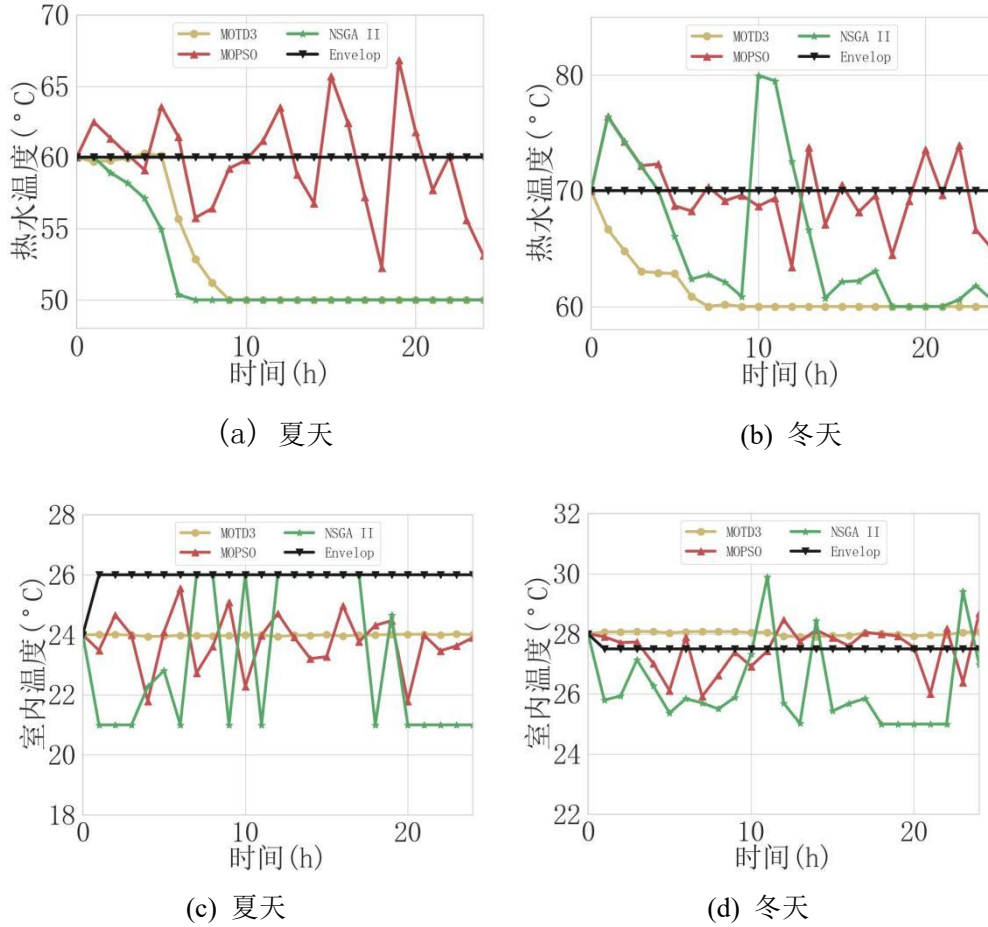


图 5-4 热水和温度对比结果

在图 5-5 中，给出了由不同算法产生的柔性载荷的时间表。当主网电价较高时，MOTD3 算法在 10:00 至 22:00 之间削减所有可削减负荷。因此，MOTD3 算法降低了运营成本和二氧化碳排放。此外，非高峰期间的可削减负荷、可移动负荷和可转移负荷基本保持不变，从而提高了客户舒适度。相比之下，其他算法的需求响应策略不能优化多个目标。

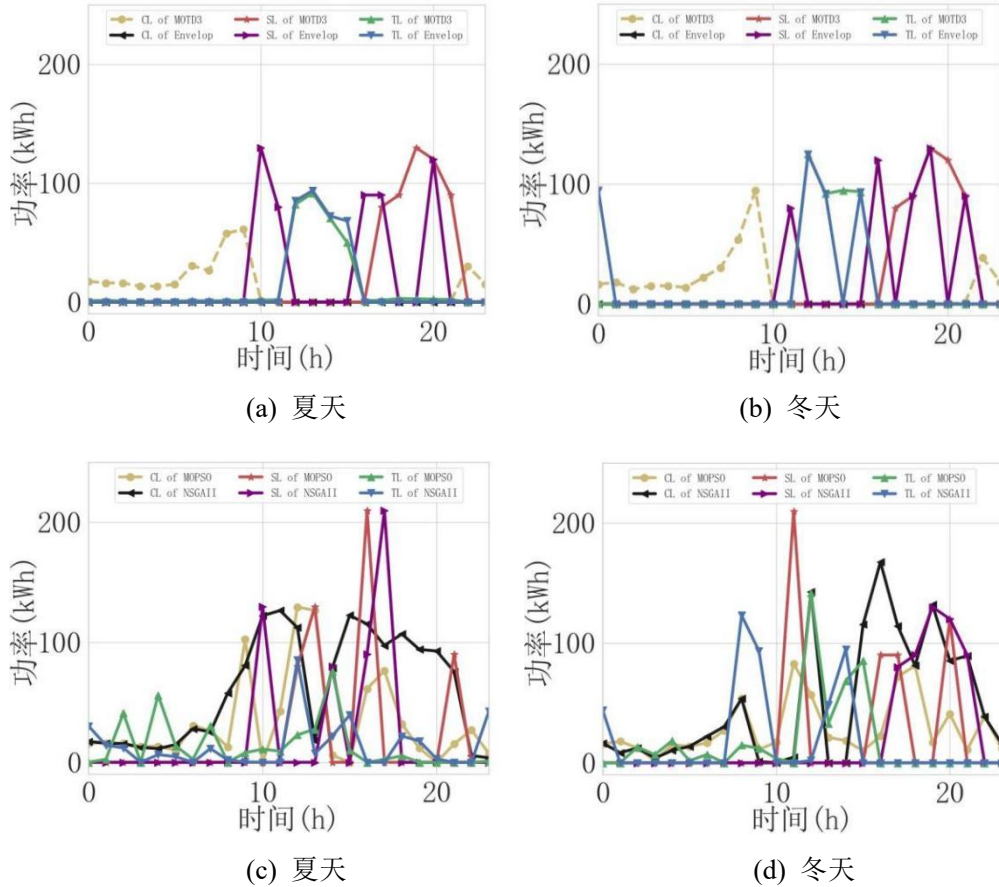


图 5-5 需求响应结果

图 5-6 显示了由各种算法生成的一天的电能调度。观察到 MOTD3 算法在非高峰电价期间储能设备储存大量电能，然后在电价相对较高且系统负荷需求较大时放电。储能设备能量调度用于减轻系统压力，并在高定价期间降低主电网的能量消耗。由于电力负荷的需求始终超过可再生能源的输出功率，MTH 和 MTC 进行大量输出，而 EB 和 EC 产生最小的输出以减少对主电网的依赖。MOTD3 算法的调度策略有效地降低了一天的运营成本和二氧化碳排放。

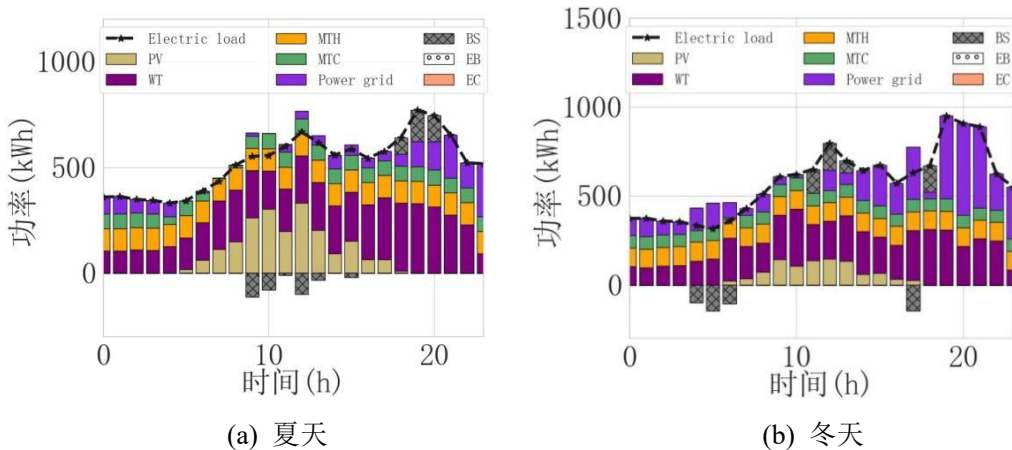


图 5-6 电能调度结果

图 5-7 显示了使用各种算法的热功率规划。很明显，MOTD3 算法主要调度 GB 和 MT 来满足加热负荷。TS 用于减轻与供热负荷相关的系统需求压力，当供热能量超过总供热需求时储存热量，当系统热能不足时释放热能。MT 在调度期间始终以最大功率运行，这可以降低 GB 的输出功率，降低系统的总运行成本。

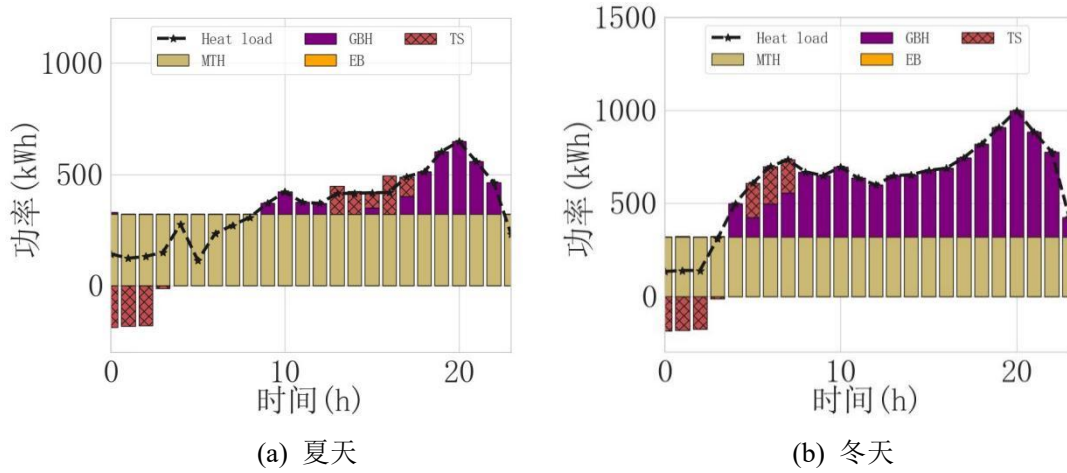


图 5-7 热能调度结果

图 5-8 显示了由不同算法生成的冷却功率时间图，我们只关注夏季冷负荷的需求。综合能源微电网的冷负荷需求相对适中，因此 MT 的输出功率提供了充足的冷却能量来满足综合能源微电网的需求，从而降低了 GB 的冷却功率需求。此外，CS 可以在系统输出功率不足时提供冷却能量，在系统冷却能量充足时储存能量。MOTD3 算法的调度策略可以降低系统的总二氧化碳排放量。

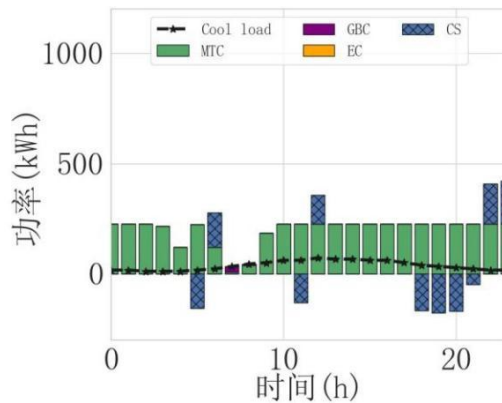


图 5-8 冷能调度结果

通过分析能量调度和需求响应调度，我们可以得出结论，我们提出的算法可以有效地同时降低运营成本、客户不满和二氧化碳排放。其他算法的调度策略不能同时优化各自目标权重下的多目标值。

表 5.7 显示了三个目标在一天的折衷调度结果。所提出的算法得到的总运行

成本、总不满意度和排放量低于其他方法。尽管在冬季，MOPSO 算法的总客户不满意度比 MOTD3 算法低 5.17%，但 MOTD3 算法的运行成本和二氧化碳排放量比 MOPSO 算法低 25.58%和 32.25%。

表 5.7 调度策略仿真结果对比

季节	目标函数	MOTD3	MOPSO	NSGA- II	Envelop
夏天	运行成本	3052.54	4298.56	5969.40	3887.86
	用户不满意度	4.41	5.29	7.92	7.07
	污染排放	4066.23	7162.30	10722.82	5519.94
冬天	运行成本	4899.84	5132.82	6555.54	5454.27
	用户不满意度	4.83	4.58	8.58	8.85
	污染排放	7922.92	11695.11	12296.84	9316.76

5.4 本章小结

在本章中，我们研究了多目标微电网综合能源调度优化问题，它被构建为多目标马尔可夫决策过程。我们提出基于 MOTD3 多目标强化学习的综合能源微电网调度优化策略，该策略可以实现对应目标权重下的最优多目标策略，通过模糊决策方法选取最终折衷策略。仿真结果表明，与现有调度优化策略相比，该策略能够最大程度减少系统成本，用户不满意，污染排放多个目标。

第 6 章 总结与展望

6.1 总结

微电网系统优化运行是保证电网经济、高效运行的重要手段。目前，为应对分布式可再生能源的随机性接入带来的影响，已有大量研究在微电网中应用人工智能的方法处理调度优化问题。本文鉴于人工智能的微电网调度优化大量研究基础之上，对基于强化学习微电网调度优化展开研究，考虑设备和系统功率平衡等约束调节，利用强化学习方法对微电网储能系统调度、多微电网协同调度优化机制、综合能源调度优化策略进行研究，以提升系统运行效益。主要结论总结如下：

1) 针对基于深度强化学习微电网系统调度优化下的行为价值评估偏差的问题，研究微电网中的储能调度规则，将微电网储能调度优化问题转化为马尔可夫决策过程。进而提出一种基于改进的深度 Q 网络方法，建立基于深度强化学习的电池充放电调度策略。仿真结果表明，与现有的强化学习充放电决策模型相比，该策略可以更有效降低了运行成本。

2) 针对多微电网运行过程中电池储能荷电状态信息缺失，加剧系统最优能源调度困难的问题，研究了非完全观测状态体下多微电网系统低功耗下的能源调度，提出了 RMADDPG 多智能体强化学习的多微电网系统协作运行机制，仿真结果表明，我们的运行机制与 DDPG、DRPG、MADDPG 和 IQL 的运行机制在非完全观测状态下相比较实现更低的系统运行功耗。

3) 针对综合能源微电网系统多目标调度优化中不同目标相互制约，导致难以优化各自对应权重下多目标的问题，本文提出了一个考虑灵活负荷和用户舒适度的能量调度问题多目标优化的新框架，该框架有三个目标：运行成本、用户不满意和二氧化碳排放。优化能量调度的目的是最大程度减小这三个目标。此外，我们将调度优化的序列决策问题转化为 MOMDP 模型。提出了一种基于 MORL 的多目标综合能源调度策略。最后引入模糊决策方法从非优势解中选择最终的折衷调度策略。仿真结果表明，我们提出的策略可以有效地降低运营成本、客户不满意度和二氧化碳排放量，并且优于其他比较算法。

6.2 未来展望

本文针对微电网能量调度与管理策略优化问题进行具体的分析和研究,对微电网的能量利用效率提升以及新能源消纳具有积极的促进作用。鉴于微电网的快速发展及新问题的不断涌现,在未来工作中,可以在以下方面进一步开展研究:

1) 由于风光发电的随机性和波动性导致微电网内部弃风弃光问题和失负荷问题严重,因此需要配置储能系统来消纳多余的风光并减少失负荷功率。传统的配置方案单纯依靠储能系统来解决以上问题,导致储能系统配置成本高、经济性较差。随着对需求侧资源的研究,将需求侧资源纳入广义储能的范围,发挥需求侧灵活充放电的特性来减少传统储能的配置成本是未来探索的方向。

2) 本文所研究的配电网系统优化与控制算法都是完全基于历史经验数据训练获得的运行优化策略,训练过程难以保证算法的完全可靠性,因此研究如何将现有基于数学机理模型的先进控制方法与基于强化学习的优化控制算法有机结合,进一步提升对复杂配电网系统的控制优化性能值得进行进一步的深入研究。

参考文献

- [1] H. Hu, N. Xie, D. Fang, X. Zhang. The role of renewable energy consumption and commercial services trade in carbon dioxide reduction: Evidence from 25 developing countries[J]. *Applied Energy*, 2018, 211(1): 1229-1244.
- [2] A. Hussain, V. Bui, H. Kim. Microgrids as a resilience resource and strategies used by microgrids for enhancing resilience[J]. *Applied Energy*, 2019, 240(15): 56-72.
- [3] G. Zhu, Y. Gao. Multi-objective optimal scheduling of an integrated energy system under the multi-time scale ladder-type carbon trading mechanism[J]. *Journal of Cleaner Production*, 2023, 417(10): 137922.
- [4] 王玉梅, 周永鑫, 张继钦. 计及需求侧资源和碳捕集的电网低碳经济调度[J]. *电工技术*, 2023, 16: 78-82.
- [5] 葛磊蛟, 范延赫, 来金钢. 面向低碳经济的人工智能赋能微电网优化运行技术[J]. *高电压技术*, 2023, 49(06): 2219-2238.
- [6] 戚陈陈, 陈鹏伟, 陈新, 等. 分散式风电高渗透率接入直流受端电网频率特性建模与分析[J]. *电网技术*, 2022, 46(06): 2161-2170.
- [7] V. Mnih, K. Kavukcuoglu, D. Silver, et al. Human-level control through deep reinforcement learning[J]. *Nature*, 2015, 518(7540): 529-533.
- [8] H. Li. Deep learning for natural language processing: advantages and challenges[J]. *National Science Review*, 2018, 5(1): 24-26.
- [9] Y. Deng, F. Bao, Y. Kong, et al. Deep direct reinforcement learning for financial signal representation and trading[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(3): 653-664.
- [10] C. Liu, H. Zhang, M. Shahidehpour, et al. A Two-Layer Model for Microgrid Real-Time Scheduling Using Approximate Future Cost Function[J]. *IEEE Transactions on Power Systems*, 2022, 37(2): 1264-1273.
- [11] E. A. M. Cese na, P. Mancarella. Energy systems integration in smart districts: robust optimization of multi-energy flows in integrated electricity heat and gas networks[J]. *IEEE Transactions on Smart Grid*, 2019, 10(1): 1122-1131 .
- [12] 易文飞, 朱卫平, 郑明忠. 计及数据中心和风电不确定性的微电网经济调度[J]. *中国电力*, 2024, 57(02): 19-26.
- [13] 崔明勇, 王楚通, 陈钰婷. 低碳环境下微电网供需联合优化调度[J]. *电工电能新技术*, 2017, 36(12): 30-38.
- [14] A. A. Eladl, M. I. El-Afifi, M. A. Saeed and M. M. El-Saadawi. Optimal operation of energy hubs integrated with renewable energy sources and storage devices considering CO₂ emissions[J]. *Int. J. Electr. Power Energy Syst.*, 2020, 117: 105719 .
- [15] X. Lu, Z. Liu, L. Ma, L. Wang, K. Zhou and N. Feng. A robust optimization approach for

- optimal load dispatch of community energy hub[J]. Appl. Energy, 2020, 259: 114195.
- [16] 朱兰, 周雪莹, 唐陇军, 劳长石. 计及可中断负荷的微电网多目标优化运行[J]. 电网技术, 2017, 41(06): 1847-1854.
- [17] A. Takeuchi, T. Hayashi, Y. Nozaki. Optimal Scheduling Using Metaheuristics for Energy Networks[J]. IEEE Transactions on Smart Grid, 2012, 3(2): 968-974.
- [18] D. EOlivares , C A. Canizares, M. Kazerani. A Centralized Energy Management System for Isolated Microgrids[J]. IEEE Transactions on Smart Grid, 2014, 5(4):1864-1875.
- [19] J. Lai, X. Lu, A. Monti, et al. Event-Driven Distributed Active and Reactive Power Dispatch for CCVSI-Based Distributed Generators in AC Microgrids[J]. IEEE Transactions on Industry Applications, 2020, 56(03): 3125-3136.
- [20] 杨军伟, 杜露露, 刘夏, 等. 高风电渗透率下考虑需求侧管理策略的智能微电网调度方法[J]. 智慧电力, 2021, 49(03): 32-39+110.
- [21] F N. Budiman, M A.M. Ramli, A H. Milyani, et al. Stochastic optimization for the scheduling of a grid-connected microgrid with a hybrid energy storage system considering multiple uncertainties[J/OL]. Energy Reports, 2022, doi.org/10.1016/j.egyr.2022.05.249.
- [22] 丁倩, 曾平良, 孙轶恺, 等. 一种考虑可再生能源不确定的分布式储能电站选址定容规划方法[J]. 储能科学与技术, 2020, 9(01): 162-169.
- [23] 严海波, 康林贤, 周东. 考虑随机性的微电网日前调度与储能优化模型[J]. 电网与清洁能源, 2019, 35(11): 61-65.
- [24] J. Tobajas, F. Garcia-Torres, P. Roncero-Sánchez, et al. Resilience-oriented schedule of microgrids with hybrid energy storage system using model predictive control[J]. Applied Energy, 2022, 306(15): 118092.
- [25] 方仍存, 桑子夏, 等. 基于改进协同量子粒子群算法的多微网负荷频率控制[J]. 电力建设, 2023, 44(07): 87-97.
- [26] 侯慧, 王晴, 等. 计及源荷不确定性及需求响应的离网型微电网两阶段日前经济调度[J]. 电力系统保护与控制, 2023, 50(13): 73-58.
- [27] 阚圣钧, 魏少雄, 等. 考虑分时电价的微电网储能容量优化配置[J]. 分布式能源, 2022, 7(02): 44-49.
- [28] 陈锦涛, 杨苹, 陈滢, 等. 基于综合需求侧响应策略的园区多能源系统优化运行[J]. 可再生能源, 2021, 39(02): 222-228.
- [29] 武昭原, 周明, 姚尚润, 等. 基于合作博弈论的风储联合参与现货市场优化运行策略[J]. 电网技术, 2019, 43(08): 2815-2824.
- [30] 张海静, 杨雍琦, 等. 计及需求响应的区域综合能源系统双层优化调度策略[J]. 中国电力, 2021, 54(04): 141-150.
- [31] X. Liu, X. Li, J. Tian, G. Yang, et al. Low-carbon economic dispatch of integrated electricity-gas energy system considering carbon capture, utilization and storage[J]. IEEE

- Access, 2023, 11: 25077 – 25089.
- [32] Z. Yang, J. Hu, X. Ai, J. Wu, et al. Transactive energy supported economic operation for multi-energy complementary microgrids[J]. IEEE Transactions on Smart Grid, 2021, 12(1): 4-17.
- [33] B. Wang, C. Zhang, C. Li, G. Yang, et al. Transactive energy sharing in a microgrid via an enhanced distributed adaptive robust optimization approach[J]. IEEE Transactions on Smart Grid, 2022, 13(3): 2279 – 2293.
- [34] 高倩, 杨知方, 李文沅. 电力系统混合整数线性规划问题的运筹决策关键技术综述与展望[J/OL]. 电工技术学报, 2023, doi:10.19595/j.cnki.1000-6753.tces.230478.
- [35] 谢旭, 张哲, 等. 基于深度强化学习的微电网内多侧储能协同调度方法[J]. 可再生能源, 2023, 41(10): 1408-1413.
- [36] 郭方洪, 何通, 等. 基于分布式深度强化学习的微电网实时优化调度[J]. 控制理论与应用, 2022, 39(10): 1881-1889.
- [37] 李咸善, 陈奥博, 等. 基于生态博弈的含云储能微电网多智能体协调优化调度[J]. 中国电力, 2021, 54(07): 166-177.
- [38] H. Wang and B. Zhang. Energy storage arbitrage in real-time markets via reinforcement learning[C]. Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM), 2018: 1-5.
- [39] 张自东, 邱才明, 张东霞, 等. 基于深度强化学习的微电网复合储能协调控制方法[J]. 电网技术, 2019, 43(06):1914-1921.
- [40] K. Zhou, K. Zhou, S. Yang. Reinforcement learning-based scheduling strategy for energy storage in microgrid[J]. Journal of Energy Storage, 2022, 51:104379.
- [41] Y. Du, F. Li. Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning[J]. IEEE Trans. on Smart Grid, 2020, 11(2):1066-1076.
- [42] H. Nunna, A. Sesetti, and A. Rathore, et al. Multiagent-based energy trading platform for energy storage systems in distribution systems with interconnected microgrids[J]. IEEE Transactions on Industry Applications, 2020, 56(3): 3207-3217.
- [43] 马冲冲, 王一铮, 等. 考虑源荷不确定性下微电网能量调度的深度强化学习策略[J]. 高技术通讯, 2023, 33(01): 79-87.
- [44] 刘向杰, 刘铎安, 等. 基于深度 Q 网络的风光柴储微电网能量管理策略[J]. 控制工程, 2023, 30(08): 1538-1547.
- [45] V. Bui; A. Hussain; H. Kim. Double deep Q -learning-based distributed operation of battery energy storage system considering uncertainties[J]. IEEE Transactions on Smart Grid, 2020, 11(1): 457-469.
- [46] L. Yu, W. Xie, D. Xie et al. Deep reinforcement learning for smart home energy management[J]. IEEE Internet Things J, 2020, 7(4): 2751-2762.
- [47] Z. Liu, Y. Liu, H. Xu. Dynamic economic dispatch of power system based on DDPG

- p>algorithm[J]. Energy Reports, 2022, 8(5):1122-1129.
- [48] 龚锦霞, 刘艳敏. 基于深度确定策略梯度算法的主动配电网协调优化[J]. 电力系统自动化, 2020, 44(06): 113-120.
- [49] 李瑜, 张占强, 孟克其劳, 等. 基于改进深度确定性策略梯度算法的微电网能量优化调度[J]. 电子测量技术, 2023, 46(02): 73-80.
- [50] X. He, S. Ge, H. Liu, et al. Frequency regulation of multi-microgrid with shared energy storage based on deep reinforcement learning[J]. Electric Power Systems Research, 2023, 214(15): 108962.
- [51] Y. Xia, Y. Xu, Y. Wang. A safe policy learning-based method for decentralized and economic frequency control in isolated networked microgrid systems[J]. IEEE Transactions on Sustainable Energy, 2022, 13(4): 1982-1993.
- [52] 刘林鹏, 朱建全, 陈嘉俊. 基于柔性策略-评价网络的微电网源储协同优化调度策略[J]. 电力自动化设备, 2022, 42(01): 79-85.
- [53] K B. Gassi, M. Baysal. Improving real-time energy decision-making model with an actor-critic agent in modern microgrids with energy storage devices[J]. Energy, 2023, 263(15): 126105.
- [54] Y. Wang, Y. Cui, Y. Li, et al. Collaborative optimization of multi-microgrids system with shared energy storage based on multi-agent stochastic game and reinforcement learning[J]. Energy, 2023, 280(1): 128182.
- [55] D. Qiu, T. Chen, G. Strbac. Coordination for multi energy microgrids using multiagent reinforcement learning[J]. IEEE Transactions on Industrial Informatics, 2023, 19(4): 5689-5700.
- [56] 郭方洪, 何通, 吴祥. 基于分布式深度强化学习的微电网实时优化调度[J]. 控制理论与应用, 2022, 39(10): 1881-1889.
- [57] 彭春华, 陈思畏, 等. 综合能源系统混合时间尺度多目标强化学习低碳经济调度[J]. 电网技术, 2022, 46(02): 4914-4925.
- [58] S.-J. Chen, W.-Y. Chiu, W.-J. Liu. User preference-based demand response for smart home energy management using multi objective reinforcement learning[J]. IEEE Access, 2021, 9: 161627 – 161637.
- [59] J. Wu, D. Song, X. Zhang, C. Duan, et al. Multi-objective reinforcement learning-based energy management for fuel cell vehicles considering lifecycle costs[J]. International Journal of Hydrogen Energy, 2023, 48(95): 37385-37401.
- [60] J. Yang, J. Liu, Y. Xiang, et al. Data-driven optimal dynamic dispatch for hydro-pv-phs integrated power systems using deep reinforcement learning approach[J]. CSEE Journal of Power and Energy Systems, 2023, 9(3): 846-858.
- [61] 李远征, 龙信鑫, 周纯杰, 等. 高比例新能源电网-分布式数据中心集群协同优化运行研究[J]. 中国科学: 技术科学, 2024, 54(01): 119-135.

- [62] E. L. Thorndike. A proof of the law of effect[J]. Science, 1933, 77(1989): 173-175.
- [63] M.L. Minsky. Theory of neural-analog reinforcement systems and its application to the brain-model problem[J]. Princeton N.J. 1954.
- [64] R. Bellman. Dynamic programming[J]. Science, 1966, 153(3731): 34-37.
- [65] S. Mahadevan, M. Maggioni. Proto-value functions: A Laplacian framework for learning representation and control in Markov decision processes[J]. Journal of Machine Learning Research, 2007, 8: 2169-2231.
- [66] S. Trevino-Martinez, R. Sawhney, C. Sims. Energy-carbon neutrality optimization in production scheduling via solar net metering[J]. Journal of Cleaner Production, 2022, 30(1): 134627.
- [67] Z. Ma, F. Dong, J. Wang, et al. Optimal design of a novel hybrid renewable energy CCHP system considering long and short-term benefits[J]. Renew Energy, 2023, 206:72–85.
- [68] T. T. Nguyen, S. K. Gupta, S. Venkatesh. Distributional reinforcement learning with maximum mean discrepancy[J]. 2020, arXiv: 2007.12354v1.
- [69] M. Volodymyr, K. Koray, S. David, et al. Playing Atari with Deep Reinforcement Learning[J]. Computing Research Repository, 2013, arXiv: 1312.5602.
- [70] G. Mehmet. Dynamic Path Planning via Dueling Double Deep Q-Network (D3QN) with Prioritized Experience Replay[J]. Applied Soft Computing, 2024, 158: 111503.
- [71] K. Li, N. Ye, S. Li, H. Wang, and C. Zhang. Distributed collaborative operation strategies in multi-agent integrated energy system considering integrated demand response based on game theory[J]. Energy, 2023, 273: 127137.
- [72] Solcast api toolkit, weather dataset. 2015.
- [73] R. E. Wang, M. Everett, J. P. How. R-MADDPG for Partially Observable Environments and Limited Communication[J]. Computer Science, 2020, arXiv: 2002.06684.
- [74] J. Guo, C. Chen, Z. Lu, et al. Wind-solar power consumption strategy based on neural network prediction and demand-side response[J]. 2020 Asia Energy and Electrical Engineering Symposium (AEEES), 2020, 29-31.
- [75] R. Lowe, Y. Wu, A. Tamar, et al. Multi-Agent Actor-Critic for Mixed Cooperative -Competitive Environments[J]. Computer Science, 2020, arXiv:1706.02275.
- [76] L. Lin, X. Guan, B. Hu, et al. Deep reinforcement learning and LSTM for optimal renewable energy accommodation in 5G internet of energy with bad data tolerant[J]. Computer Communications, 2020, 156(15):46-53.
- [77] A. Tampuu, T. Matiisen, D. Kodelja, et al. Multiagent Cooperation and Competition with Deep Reinforcement Learning[J]. Computer Science, 2015, arXiv:1511.08779.
- [78] T. Ma, J. Wu., and L. Hao. Energy flow calculation and integrated simulation of micro-energy grid with combined cooling, heating and power[J]. Automat. Electric Power Syst., 2016, 40(23): 22-27.
- [79] F. Brahman, M. Honarmand, and S. Jadid. Optimal electrical and thermal energy management

- of a residential energy hub, integrating demand response and energy storage system[J]. *Energy Buildings*, 2015, 90: 65-75.
- [80] R. Liu, T. Yang, G. Sun, S. Lin, F. Li, and X. Wang. Multi objective optimal scheduling of community integrated energy system considering comprehensive customer dissatisfaction model[J]. *IEEE Transactions on Power Systems*, 2023, 38(5): 4328-4340.
- [81] S. Gheouany, H. Ouadi, F. Giri, et al. Experimental validation of multi-stage optimal energy management for a smart microgrid system under forecasting uncertainties[J]. *Energy Conversion and Management*, 2023, 291(1): 117309.
- [82] Z. Lu, Y. Gao, C. Xu, et al. Configuration optimization of an off-grid multi-energy microgrid based on modified NSGA-II and order relation-TODIM considering uncertainties of renewable energy and load[J]. *Journal of Cleaner Production*, 2023, 383(10): 135312.
- [83] R. Yang, X. Sun, and K. Narasimhan. A generalized algorithm for multi-objective reinforcement learning and policy adaptation[J]. In *Advances in Neural Information Processing Systems*, 2019, arXiv:1908.08342.

攻读学位期间发表的学术成果

- [1] **Jiakai Gong**, Nuo Yu*, Fen Han, Bin Tang, Yuan Ge, Scheduling strategy for energy storage system in microgrids employing deep reinforcement learning[C]. 2023 China Automation Congress (CAC), Chongqing, China, 2023: doi: 10.1109/CAC59555.2023.10450967.
- [2] Fen Han, Nuo Yu*, **Jiakai Gong**, Yuan Ge and X. Gao, Task scheduling for mobile edge computing leveraging deep reinforcement learning[C]. 2023 42nd Chinese Control Conference (CCC), Tianjin, China, 2023. doi: 10.23919/CCC58697.2023.10240306.
- [3] 申请相关发明专利一项：一种基于值分布深度 Q 网络的微电网储能调度优化方法（申请号：2023108704543）。

致 谢

首先，我要对我的导师表达衷心的感谢之情。完成这篇论文的过程中，导师予以了精心的心理引导，从选定论文主题、选择和汇总相关文献资料，到论文的修订等方方面面，导师耗费了大量的时间和精力，给予了我宝贵的指导和支持。此外，导师严谨负责的态度和治学精神也对我产生了深远的影响，我由衷地感谢导师的关怀和悉心指导。

其次，我要感谢涉及到本文的各位学者。在本文中引用了多位学者的研究文献，没有他们的研究成果和启发，我将难以完成这篇论文的撰写工作。

最后，我要衷心感谢参与论文审阅和答辩的老师。在繁忙的工作中，他们抽出宝贵的时间对本文进行审阅，我对他们的认真态度和支持表示由衷的感激。

尚德敏学 唯实惟新

