

分类号: TM464

密 级: 公开

学校代码: 10363

学 号: 2210330158



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
网能量管理研究

论 文 作 者 蒋卓伟
指 导 教 师 娄柯
学 科 (专 业) 电子信息
研 究 方 向 控制工程与自动化技术

论文提交日期: 2024 年 05 月 31 日

分 类 号 ： TP464

密 级 ： 公 开

单位代码：10363

学 号：2210330158

题 目 基于深度强化学习的微电网能量管理的研究

题 目 Research on microgrid energy management
based on deep reinforcement learning

学生姓名： 蒋卓伟

指导教师： 娄柯

专 业： 电子信息

研究方向： 控制工程与自动化技术

论文答辩日期： 2024 年 5 月 24 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后果由本人承担。

学位论文作者签名：蒋年伟

日期： 2024 年 5 月 31 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。

本学位论文属于保密 ☐， 不保密 ☒。在 _____ 年解密后适用本版权书。

学位论文作者签名： 蒋年伟

指导教师签名： 姜柯

日期： 2024 年 5 月 31 日

日期： 2024 年 5 月 31 日

基于深度强化学习的微电网能量管理研究

摘要

微电网能量管理可以在促进可再生能源的整合的同时提升微电网的运行效率并降低运行成本。但随着微电网技术的迅速发展,使用传统方法解决微电网能量管理问题面临着适应性不足、优化能力有限、响应速度慢和成本较高等问题,使用能够处理复杂运行场景、具有自适应学习能力的方法成为微电网能量管理未来发展的方向。因此本文提出了基于深度强化学习的微电网能量管理研究。论文的主要研究内容如下:

(1)根据深度强化学习的理论框架构建了微电网的能量管理模型,并基于该模型使用深度Q网络(Deep Q-Network, DQN)算法和深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)算法实现能量管理。通过仿真实验表明,使用深度强化学习算法对微电网进行能量管理,可以实现减少运行成本的目标。

(2)针对单一微电网提出了一种基于长短期记忆神经网络和奖励优化的深度确定性策略梯度的微电网能量管理方案(Long and short-term memory neural networks and Reward Optimization Deep deterministic policy gradient algorithm, LRODDPG)。在现有基于深度强化学习能量管理方案中并未对模型环境中的不确定性因素进行单独建模,而是将微电网状态构建为状态向量直接输入神经网络,降低了模型的收敛速度。为此本文使用长短期记忆神经网络对微电网能量管理模型中的随机环境进行建模,提取源荷时序数据特征,通过训练得到有效的预测模型,降低源荷不确定性对模型的影响。同时为了进一步提高算法的收敛性,本文还对奖励函数进行了优化,将奖励函数分

为即时奖励和完成奖励，根据微电网环境状态和优化目标在训练过程中的不同阶段给与智能体不同的奖励信号反馈。最后通过对比实验，验证改进后的算法可以有效降低运行成本并提高系统自平衡率。

(3)在单一微电网能量管理的基础上，针对多微电网协同控制难度大的问题，提出一种基于多智能体深度确定性策略梯度和最优潮流的双层能量管理方案(Bilayer Multi-Agent Deep Deterministic Policy Gradient, BMADDPG)。该方案分为两层，首先下层使用更适用于多智能体环境的 MADDPG 算法对多微电网中的设备出力进行决策；随后上层根据下层智能体的优化结果使用最优潮流模型求解多微电网间的最优潮流，解决了协同控制难度大的问题，并实现了微电网间的能量互济和功率平衡。最后通过实验仿真验证了方案的有效性，并通过对比实验证明该方案可以在保证执行速度的同时，以最低成本完成能量管理。

关键词：微电网，能量管理，深度强化学习，奖励优化，最优潮流

Research on microgrid energy management based on deep reinforcement learning

ABSTRACT

Microgrid energy management can enhance the operational efficiency of microgrids and reduce operational costs while facilitating the integration of renewable energy sources. With the rapid development of microgrid technology, the use of traditional methods to solve the microgrid energy management problem faces problems such as lack of adaptability, limited optimization capability, slow response time and high cost, and the use of methods that can handle complex operational scenarios and have adaptive learning capability has become the future direction of microgrid energy management. Therefore, this paper proposes a research on microgrid energy management based on deep reinforcement learning. The main research content of this paper is as follows:

(1) An energy management model for microgrids is constructed based on the theoretical framework of deep reinforcement learning, and energy management is realized based on the model using Deep Q-Network (DQN) algorithm and Deep Deterministic Policy Gradient (DDPG) algorithm. Simulation experiments show that energy man

agement of microgrids using deep reinforcement learning algorithms can achieve the goal of reducing operating costs.

(2) Long and short-term memory neural networks and Reward optimization Deep deterministic policy gradient algorithm(LRODDPG) is proposed for a microgrid energy management scheme for a single microgrid. In the existing deep reinforcement learning based energy management schemes the uncertainties in the model environment are not modeled separately, but the microgrid states are constructed as state vectors directly input into the neural network, which greatly reduces the convergence speed of the model. For this reason, in this paper, we use a long and short-term memory neural network to model the stochastic environment in the microgrid energy management model, extract the source-load time-series data features, and obtain an effective prediction model through training to reduce the impact of source-load uncertainty on the model. In order to further improve the convergence of the algorithm, this paper also optimizes the reward function, which is divided into immediate reward and completion reward, and gives different reward signal feedbacks to the intelligent body according to the state of the microgrid environment and optimization goals at different stages in the training process. Finally, it is verified through comparative experiments that the improved algorithm can effectively reduce the operation cost and improve

ve the convergence speed.

(3) On the basis of single microgrid energy management, a bilayer energy management scheme based on multi-intelligent deep deterministic policy gradient and power flow is proposed for the problem of difficult cooperative control of multiple microgrids(Bilayer Multi-Agent Deep Deterministic Policy Gradient ,BMADDPG) . The scheme is divided into two layers, firstly, the lower layer uses the MADDPG algorithm, which is more suitable for the multi-intelligent body environment, to make decisions on the equipment output in the multi-microgrid; subsequently, the upper layer uses the optimal tidal current model to solve the optimal tidal currents among the multi-microgrids according to the optimization results of the lower layer of intelligences, which solves the problem of the difficulty of collaborative control and realizes the energy mutualization and power balance among the microgrids. Finally, the effectiveness of the scheme is verified by experimental simulation, and it is proved by comparative experiments that the scheme can accomplish energy management at the lowest cost.

KEY WORDS: Microgrid, Energy Management, Deep Reinforcement Learning, Reward Optimization, Optimal Power Flow

目录

摘要	I
ABSTRACT	III
第 1 章 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
1.3 研究内容与章节安排	5
1.3.1 主要研究内容	5
1.3.2 章节安排	6
第 2 章 相关理论	7
2.1 强化学习概念及基本原理	7
2.1.1 强化学习背景	7
2.1.2 马尔可夫决策过程	8
2.2 深度强化学习算法	10
2.2.1 基于价值函数的深度强化学习算法	10
2.2.2 基于策略梯度的深度强化学习算法	11
2.3 源荷预测模型	13
2.3.1 循环神经网络	13
2.4 本章小结	16
第 3 章 基于深度强化学习的微电网能量管理模型	17
3.1 微电网系统数学模型	17
3.2 基于深度强化学习的微电网能量管理模型	19
3.3 算例分析	21
3.3.1 仿真实验数据及参数定义	21
3.3.2 仿真结果与分析	22
3.4 本章小结	24
第 4 章 基于 LRODDPG 的单微电网能量管理	26
4.1 引言	26
4.2 单微电网能量管理模型	26

4.2.1 优化目标与奖励函数定义	26
4.2.2 状态与动作空间定义	27
4.3 基于 LRODDPG 的单微电网能量管理方案	28
4.3.1 基于 LSTM 的源荷时序数据提取	28
4.3.2 奖励函数优化设计	30
4.4 实验结果分析	33
4.4.1 仿真实验数据	33
4.4.2 训练效果对比	34
4.4.3 调度结果对比	35
4.5 本章小结	37
第 5 章 基于 BMADDPG 的多微电网双层能量管理	39
5.1 引言	39
5.2 多微电网双层能量管理模型	39
5.2.1 下层多智能体深度强化学习算法	40
5.2.2 上层最优潮流模型	41
5.3 基于 BMADDPG 的多微电网双层能量管理方案	43
5.3.1 约束条件	43
5.3.2 优化目标函数	44
5.3.3 构建马尔可夫决策过程	45
5.3.4 MADDPG 算法流程	46
5.4 实验结果分析	48
5.4.1 仿真实验数据	48
5.4.2 训练效果分析	49
5.4.3 调度结果分析	49
5.4.4 不同方案效果对比	53
5.5 本章小结	54
第 6 章 总结与展望	55
6.1 总结	55
6.2 展望	56

参考文献.....	57
攻读硕士期间的学术成果.....	63

第 1 章 绪论

1.1 研究背景及意义

随着化石能源的过度使用和工业化的快速发展,当前气候环境问题变得愈发严峻,世界各国也开始逐步开展减少温室气体排放的行动。可再生能源作为一种低碳、清洁的能源形式,被广泛认可为减少碳排放和缓解气候变化的重要手段。国际环境意识的提高促使各国政府采取了积极的政策措施,推动可再生能源的大规模利用^[1]。

随着科学技术的不断进步,可再生能源技术在效率和可靠性方面取得了巨大的突破。此外,储能技术的不断改进和成本下降,也为可再生能源的可持续利用提供了重要支持。传统能源形式如石油和天然气的供应受到地缘政治和市场价格波动的影响,给国家的能源安全带来了风险^[2]。因此,为了增强能源供应的多样化和安全性,各国政府纷纷加大可再生能源的发展和利用力度。可再生能源的广泛应用不仅能够减少能源成本,降低电力价格,还能为当地创造就业机会,促进经济增长。为了鼓励可再生能源的发展和利用,各国政府纷纷出台了一系列的政策和法规,如补贴政策、税收激励措施、可再生能源配额制度等,以推动可再生能源的市场化和商业化^[3]。这些政策和法规的实施为可再生能源的快速增长提供了重要的支持和保障。

分布式发电技术的起源可以追溯到 20 世纪初。在这个阶段,分布式发电主要以小型燃气发电机组和蓄电池组为主,用于供应电力较小规模的场所如住宅、商业建筑等^[4]。20 世纪 60 年代至 80 年代,分布式发电技术经历了一系列的技术演进。其中,燃料电池技术、微型涡轮发电机等新兴技术得到了广泛关注和研究^[5]。进入 21 世纪以来,分布式发电技术迎来了快速发展的阶段。这一阶段的关键推动因素包括可再生能源技术的成熟和商业化、能源转型的迫切需求以及政策和法规的支持^[6]。在这个阶段可再生能源技术得到了广泛应用,并成为分布式发电的主要形式。同时,人工智能、区块链等新兴技术的应用也为分布式发电的智能化、自主化提供了新的可能性^[7]。未来的分布式发电系统将更加智能、可靠和可持续,为实现清洁能源和可持续发展目标做出更大贡献。

利用深度强化学习对微电网进行能量管理是当前微电网领域的研究热点之

一。深度强化学习是一种基于深度学习和强化学习相结合的方法，相较于传统方法解决微电网能量调度问题，使用深度强化学习算法在安全性、经济性和实时决策能力方面都有着显著提升，同时还具备传统方法所不具备的自适应性和可扩展性，大大降低了人工操作的复杂性^[8]。目前，虽然深度强化学习在微电网能量管理方面的应用仍处于研究阶段，但在能源优化调度、负荷预测、储能控制和优化、智能负荷管理等领域已经取得了一些重要进展^[9]。尽管深度强化学习在微电网能量管理方面具有很大的潜力，但目前仍面临一些挑战。其中包括数据需求、训练过程的复杂性、算法调优和实时性等方面的问题^[10]。因此，未来需要进一步的研究和实践以克服这些挑战，并将深度强化学习应用于微电网能量管理的实际应用中。

1.2 国内外研究现状

微电网能量管理系统是一种旨在实现分布式能源资源优化配置与运行控制的智能化系统。其核心功能是在确保微电网安全、稳定、高效运行的前提下，对微电网内部的分布式能源（如风能、太阳能等）、储能装置及负荷进行综合管理^[11]。微电网能量管理是一个复杂的优化问题，涉及到多个变量和约束条件，需要综合考虑经济效益、运行安全、电力质量等多个方面^[12]。目前，解决微电网能量管理问题的方法主要包括以下几种：

(1) 数学优化方法：这是解决微电网能量管理问题的传统方法，主要使用线性规划、非线性规划、和动态规划等数学工具^[13]。这些方法可以处理包括约束条件和优化目标在内的多种复杂问题。Garip I 提出混合整数线性规划（MILP）方案解决微电网能量管理问题，降低了电压波动情况和有功功率损耗^[14]。Akulker H 提出了一种单层确定性混合整数非线性规划（MINLP）来设计和调度 PTG 集成微电网，将最佳设备选择、设计和调度仅在一个公式中确定，降低了计算难度^[15]。Shuai H 提出了一种基于近似动态规划（ADP）的分布式发电微电网经济调度方法，采用分段线性函数（PLF）近似和改进的斜率更新策略并蒙特卡罗方法对训练场景进行采样，为 ADPED 提供经验知识^[16]。

使用数学优化方法解决微电网能量管理问题时，虽然能够提供高效的能源分配方案但仍然存在一些不足之处：随着微电网规模的扩大和参与元素的增加，模型复杂度将显著提高，数学优化模型通常需要考虑多个变量和约束条件，对计算

能力和算法的效率提出更高要求。传统的优化算法可能在面对高度非线性、大规模问题时计算速度慢，难以在合理时间内得到满意解，同时优化算法通常需要基于预测数据进行计算，传统优化算法在实时调整能力方面表现不足^[17]。

(2)启发式算法：随着计算机技术的发展，启发式算法在微电网能量管理中的应用越来越广泛，这类算法主要包括遗传算法、粒子群优化算法、模拟退火算法等，这些算法通常不需要问题严格的数学模型，能够在合理的时间内找到满意的解决方案^[18]。Nemati M 等提出了一种改进的实编码遗传算法的增强型微电网单元承诺调度和经济调度方法，所采用的遗传算法具有高度灵活的子函数集、智能收敛行为以及多样化的搜索方法和对约束违规的惩罚方法^[19]。李国庆等提出了一种改进蚁群算法，采用信息混合更新策略的蚁群算法会首先找到更多的可行解，还能保持继续搜索最优解的能力，同时保持较快的收敛速度^[20]。Hafez A 等提出了一种简单高效的混合模拟粒子群退火算法（SAPS），该算法考虑了可再生能源的间歇性，通过调整电池容量增强电源可靠性并充当虚拟同步惯性，提高了微电网系统供电的稳定性^[21]。

启发式算法在微电网能量管理中的应用，提供了一种有效的方法来处理实际中的复杂性和不确定性，同时保持了足够的灵活性和鲁棒性。但由于在处理微电网能量管理问题会涉及多个分布式能源和储能设备，这些设备的相互作用可能导致使用启发式算法往往只能求得局部最优解而非全局最优解^[22]；同时随着微电网规模的快速增长，启发式算法在处理大规模问题时可能会遇到性能瓶颈，算法的执行时间和计算复杂度可能会随着问题规模的增加而显著增加^[23]。

(3)模型预测控制（MPC）：模型预测控制是一种先进的控制策略，它通过建立系统的动态模型，预测未来一段时间内的系统状态，并优化控制输入以达到预定的目标^[24]。MPC 在微电网能量管理中可以用来实时调整可再生能源的发电量和储能设备的充放电策略。Kong X 等构建了一种分层分布式模型预测控制（HDMPC）。在该 HDMPC 中，上层利用迭代分布式控制策略来实现电力调度的协调。下层利用监控预测控制来实现经济性和跟踪性。在这种分层结构下，利用从下层控制层到上层的反向计算来保持约束的一致性。通过各子系统之间的协同优化，实现了分布式能源的即插即用^[25]。

尽管 MPC 在微电网能量管理中展现出许多优势，但也存在一些不足之处：

MPC 需要在每个时间步长中进行优化计算，这通常涉及到求解一个非线性优化问题。对于包含多种能源和负载的复杂微电网系统，预测模型的建立和优化问题可能变得非常复杂，从而导致计算时间和计算资源的需求显著增加^[26]；同时 MPC 的效果很大程度上依赖于预测模型的准确性而微电网系统具有非线性、不确定性以及时变性等特点，这些特性可能导致预测模型难以准确描述实际系统行为，模型的不准确可能会导致控制效果的下降，甚至可能引起系统的不稳定^[27]。

(4)深度强化学习：深度强化学习技术在微电网能量管理中也发挥着重要作用，目前已经在微电网的能量管理中得到了广泛的应用,实现了对微电网的实时控制和反馈^[28]。其中，最具代表性的算法是深度 Q 网络（DQN），DQN 通过引入深度神经网络来近似 Q 函数，有效解决了高维状态空间的问题^[29]。Sun S 等提出了一种将 DQN 算法与多重注意力机制相结合的协同优化方法，该算法计算系统状态的加权特征，以计算 Q 值和优先级，以确定能源系统的下一个操作指令，为处理能源损失等问题提供了一种新的方法，有望提高微电网等领域的经济效率和可持续性^[30]。随着研究的深入，研究者提出了基于多智能体协同学习、分层决策等方法的深度强化学习算法。Gao J 等为了增强处理各种不确定性的泛化能力，提出了一种改进的多智能体 SAC 算法（MASAC），该算法促进了多微电网中多智能体之间的能量交易，并采用自动化机器学习对 MASAC 超参数进行优化，以进一步提高深度强化学习的泛化性^[31]。Li J 等为了解决孤立多区域微电网的协调问题，提出了一种群体智能负载频率控制（SI-LFC）方法，将各区域的微电网等是为独立智能体，并采用群体智能集中离线学习策略，实现利益平衡^[32]。Jing Y 等提出一种基于多智能体深度强化学习的两级协同控制策略模型。该模型的上层是基于多智能体深度确定性策略梯度算法（MADDPG）的多智能体混合博弈决策模型，下层包含功率和气流计算模型，下层模型为上层模型提供 IES 的能量流数据，上层根据下层提供的能量流数据奖励智能体的决策行为，有效求解复杂耦合网络中存在的高维非线性优化问题，该方法可以提高模型的收敛性和训练速度^[33]。

目前应用于微电网能量调度的深度强化学习算法主要有以下几种：DQN 算法，DQN 算法是最早广泛被应用于能量管理的深度强化学习算法之一，能够处理离散空间动作，适用于动作空间有限的场景，这也意味着在面对复杂微电网

环境时该算法具有一定的局限性^[34]；异步优势演员评论家(A3C)算法,A3C 算法采用异步更新策略,可以让多个智能体在不同节点上并行学习,然后将学到的策略和值函数汇总起来,适用于需要连续控制的情况,如多个分布式能源的调度^[35]；近端策略优化(PPO)算法, PPO 算法通过近端策略梯度的方法来改进传统的策略梯度算法,将复杂的策略更新问题转化为一系列简单的近端优化问题,减少微电网能量管理中由于决策不确定性带来的风险,适用于需要快速收敛和稳定性的场景^[36]；深度确定性策略梯度(DDPG)算法,与传统的策略梯度算法相比,DDPG 算法采用确定性策略,即在给定状态下选择一个固定的动作,可以降低微电网能量管理中的不确定性,特别适合于那些需要快速反应和连续控制的场景^[37]。

1.3 研究内容与章节安排

1.3.1 主要研究内容

本文针对微电网能量管理问题提出了基于深度强化学习的能量管理方案。首先介绍了相关的深度强化学习算法以及能量管理方案,其次对单一微电网和多微电网环境下的能量管理分别进行了研究。论文的主要研究内容如下:

(1) 建立微电网的数学模型,并根据深度强化学习的理论框架构建了微电网的能量管理模型,并基于该模型使用 DQN 算法和 DDPG 算法实现能量管理。

(2) 针对单一微电网,本文提出了基于 LRODDPG 的单微电网能量管理方案。为了解决在传统深度强化学习能量管理方案中由于可再生能源不确定导致的算法收敛较差的问题,本文使用长短期记忆神经网络对微电网能量管理模型中的随机环境进行建模,提取源荷时序数据特征,通过训练得到有效的预测模型,降低源荷不确定性对模型的影响。此外针对奖励函数进行了重新设计,将奖励函数分为即时奖励和完成奖励,对训练过程中的不同阶段给与智能体不同的奖励信号反馈,提高了算法的收敛性。

(3) 针对多微电网,本文提出了基于 BMADDPG 的多微电网双层能量管理方案。多微电网环境下的系统整体的拓扑结构和运行状态更加复杂,需要协调各个微电网之间的能源流动和负载分配,因此多微电网能量管理需要更为先进的管理策略和控制算法。针对上述问题本文在处理多微电网能量管理问题时,提出基于 BMADDPG 的多微电网双层能量管理方案,首先上层使用多智能体深度强化学

习算法对微电网的设备出力进行决策，随后下层将决策的结果进行最优化潮流，保证多微电网间的功率平衡的同时增强了系统稳定性和决策的精确性。

1.3.2 章节安排

本文分为六章，其结构安排如下：

第一章首先对微电网的发展历程进行了概括，其次对当前微电网能源管理方案进行综述，最后阐述了本文的主要研究内容。

第二章介绍了深度强化算法的原理和相关基础知识，对本文所使用的算法进行了详细介绍为后文实现微电网能量管理方案做好铺垫。

第三章搭建微电网的数学模型，并在深度强化学习框架下建立了微电网能量管理模型，最后使用深度强化学习算法在模型上训练智能体学习能量管理策略，实现基于深度强化学习的微电网能量管理。

第四章基于 LRODDPG 的单微电网能量管理方案。在 DDPG 算法的基础上加入长短期记忆神经网络提取源荷时序数据特征，并对奖励函数进行重新设计，最后通过实验分析将改进后的算法与原算法进行对比，验证方案的有效性。

第五章基于 BMADDPG 的多微电网双层能量管理方案。首先构建了多微电网的双层能量管理模型，包含最优潮流模型和多智能体深度强化学习模型，最后通过实验仿真证明该方案的有效性，并与其他方案进行对比验证该方案的优越性。

第六章为总结与展望。对本文的主要研究内容进行总结，指出了当前研究内容的不足之处，并对未来深度强化学习与微电网的深度结合做出了展望。

第 2 章 相关理论

2.1 强化学习概念及基本原理

2.1.1 强化学习背景

深度强化学习融合了深度神经网络、算法自动化及控制理论等多领域理念，成为探究智能决策的关键途径，同时也是通往通用人工智能的一条途径。Zhu Z 等将强化学习阐述为：智能体通过试错方法，学会在不同的状态中最优地选择动作以实现长期累积奖励的最大化^[38]。深度强化学习把深度学习的感知功能与强化学习的决策功能融为一体。通过从输入数据中提取特征，进而实现对系统的控制。这种方法模拟了人类的思考方式，展现出强大的能力^[39]。如图2-1所示，深度强化学习的基本架构中，智能体通过感知环境状态并采取相应的动作，动作是智能体对环境状态的响应，它可以改变环境状态。环境根据智能体的动作给出反馈，即奖励信号，智能体依据奖励信号调整自己的行为策略，最大化累积奖励。神经网络会根据累计奖励对参数进行调整，不断进行学习获得最高奖励。通常来说，一项高新技术的提出最终需要实际应用来予以实现。期望智能体能够替代人类进行决策，例如，根据实际情况智能体可以控制机械臂抓取物品。在这个场景下，物品的摆放就构成了智能体所处的环境。智能体的行为包括移动路径、水平或垂直位置的调整；能否在确保安全的基础上迅速提取物品等行为可视为奖励标准。从上述示例中可以看出对实际场景中的智能体与环境互动的研究才是具有实际意义的。在机器学习领域，这种交互过程称为序列决策（sequential decision making）^[40]。

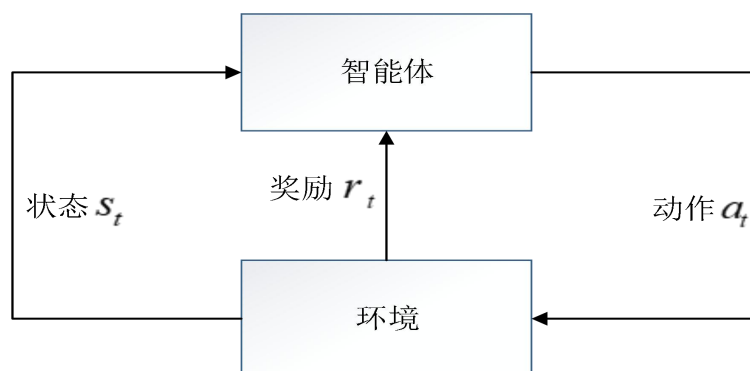


图 2-1 强化学习的总体框架

2.1.2 马尔可夫决策过程

深度强化学习主要包括四个方面：智能体、智能体的环境、执行的动作以及获得的奖励。在未知环境中，智能体首先将对环境进行探索获得状态信息，随后根据获得的信息执行动作，智能体也会获得相应奖励作为反馈，最后获得奖励后的智能体会更新策略执行新的动作。智能体通过与环境持续互动并接收奖励反馈，不断调整动作策略，最终掌握能带来最高累计回报的操作^[41]。深度强化学习框架可以通过一个4元组 $\langle S, A, P, R \rangle$ 和策略 π 构建马尔可夫决策过程进行表述。智能体与环境的一次交互可以抽象化为：在时刻 t ，环境处于状态 $S_t \in S$ ，智能体根据策略 π 采取一个动作 $A_t \in A$ ，获得一个即时奖励 $R_t \in R$ ，然后环境的状态转移到 $S_{t+1} \in S$ 。

(1)状态空间 $S = \{s_1, s_2, s_3 \dots s_n\}$ ：由智能体所处的环境组成， s_i 代表系统可能处于的每一个状态。

(2)动作空间 $A = \{a_1, a_2, a_3 \dots a_n\}$ ：动作空间的大小和选择的动作将直接影响智能体的决策质量和性能。

(3)状态转移函数 $[P(s, a, s') = P(s' | s, a)]$ ：其中 $P(s' | s, a)$ 给出了在状态 s 采取动作 a 后，系统转移到状态 s' 的概率。

(4)奖励函数 R ：奖励函数定义了在一定状态下执行动作所能获得的即时奖励，一般形式为 $r(s, a, s')$ ，这个函数对每个可能的环境动作返回一个实数，表示在状态 s 执行动作 a 后，智能体获得的奖励。

(5)智能体采取的策略 π ：是指在每个状态下选择动作的概率分布策略可以表示为 $\pi(a | s)$ 表示在状态 s 下选择动作 a 的概率。

在智能体在与环境的互动中的目标是最大化长期累积的收益。如前述智能体操控机械手臂的实例，若智能体仅关注眼前收益，过度侧重于当前时刻的决策，例如迅速抓取，虽然短期效益较高，但可能因动作过快导致物品破损。因此，为应对所述问题，引入了累积奖励。在每次迭代中，智能体执行操作，改变环境状态并获取奖励。定义累计奖励为 R_t ：

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \quad (2-1)$$

上式中 $\gamma \in (0,1)$ 为衰减因子，当 γ 值设得越靠近零时，智能体就越是偏向于追求短期累积奖励。相反，当 γ 值越靠近 1 时，智能体就越偏好获取长期累积的奖励。

为了评估智能体探索的状态质量高低，引入状态价值函数 $V^\pi(s)$ 它表示在某个状态 s 下，智能体预期能够获得的长期累积奖励。它可以帮助智能体评估状态的好坏，并指导智能体选择最优的动作 a 。状态价值函数通常通过神经网络进行估计，可以表示为：

$$\begin{aligned} V^\pi(s) &= E_\pi[R_t | S_t = s] \\ &= E_\pi \left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s \right] \end{aligned} \quad (2-2)$$

同理，为了评估智能体执行动作的质量优劣，引入动作价值函数 $Q^\pi(s, a)$ ：

$$\begin{aligned} Q^\pi(s, a) &= E[R_t \mid s_t = s, a_t = a] \\ &= E \left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a \right] \end{aligned} \quad (2-3)$$

状态价值函数 $V^\pi(s)$ 是动作价值函数 $Q^\pi(s, a)$ 关于动作 a 的期望；动作函数 $Q^\pi(s, a)$ 是状态价值函数 $V^\pi(s)$ 与动作奖励反馈值的和的期望，即：

$$Q^\pi(s, a) = E[r_{t+1} + \gamma V^\pi(s_{t+1})] \quad (2-4)$$

$$V^\pi(s) = E_a[Q^\pi(s, a)] \quad (2-5)$$

依据贝尔曼方程(Bellman Equation)，价值函数呈现递推特性，具体推导如下：

$$\begin{aligned} V^\pi(s_t) &= E[R_t \mid s_t = s] \\ &= E[r_{t+1} + \gamma r_{t+2} + \dots \mid s_t = s] \\ &= E[r_{t+1} + \gamma(r_{t+2} + \gamma r_{t+3} + \dots) \mid s_t = s] \\ &= E[r_{t+1} + \gamma V^\pi(s_{t+1}) \mid s_t = s] \end{aligned} \quad (2-6)$$

动作价值函数也可以通过上式进行类似推导，最终表达式为：

$$Q^\pi(s_t, a_t) = E[r_{t+1} + \gamma Q^\pi(s_{t+1}, a_{t+1}) \mid s_t = s, a_t = a] \quad (2-7)$$

价值函数是用来评估策略 π 的，当策略在特定约束下，如状态空间和动作空间呈离散有限状态时，可以通过对这些有限策略进行评价，进而选出最佳策略，

即：

$$\pi^* = \arg \max_{\pi} V^{\pi}(s) \quad (2-8)$$

$$\pi^* = \arg \max_{\pi} Q^{\pi}(s, a) \quad (2-9)$$

当状态空间和动作空间连续时，无法通过有限的策略搜索得到最优策略，可以通过将值函数表示为策略参数的某个函数，进而求出值函数关于策略参数的梯度，并使参数沿着梯度上升的方向更新，以实现目标函数 $J(\theta)$ 最大化，其中：

$$J(\theta) = \mathbb{E}[R_t] = \int_{\tau \sim \pi_{\theta}(\tau)} \pi_{\theta}(\tau) R_{\tau} d\tau \quad (2-10)$$

θ 表示特征参数， π_{θ} 可视做回报的概率密度函数。

2.2 深度强化学习算法

2.2.1 基于价值函数的深度强化学习算法

基于价值函数的深度强化学习算法是一种将深度学习技术应用于强化学习问题的方法。该算法主要通过构建价值函数来评估和优化策略，使得智能体能够在与环境交互的过程中获得最大的累积奖励。

该算法的核心思想是利用深度神经网络来表示价值函数，并通过与环境的交互来更新网络的参数。在训练过程中，算法会根据当前状态和采取的动作，计算预测的奖励值与实际奖励值之间的误差，并使用反向传播算法来更新网络的权重。

其中深度 Q 网络(DQN)是一种典型方法。与 Q 学习相较，DQN 主要做出三点改进：1 使用了深度神经网络来逼近 Q 值函数，能够处理高维度的状态空间和动作空间。2 使用了经验缓存技术，能够存储和重用历史经验，减少样本相关性和数据稀疏性的问题。3 使用了固定目标网络来稳定训练过程，避免过拟合。4 使用了双网络结构来提高估计 Q 值的准确性。

算法 2-1 DQN 算法^[42]

输入：状态空间 S，动作空间 A

1: 初始化经验缓存 D

2: 随机初始化 Q 网络参数 θ 以及目标 Q 网络参数 θ^-

3: for $i = 0 \rightarrow M$ do

4: 初始化起始状态 S

```

5:   for t = 0 → T do
6:       当处在状态 S 时，根据  $\varepsilon$  概率选择动作  $a$ 
7:       执行动作  $a$ ，得到奖励值  $r_{t+1}$  和新的状态  $s_{t+1}$ 
8:       将  $\{s_j, a_t, r_t, s_{t+1}\}$  存入  $D$  中
9:       从  $D$  中采样一批样本  $\{s_j, a_t, r_t, s_{t+1}\}$ 
10:       $y_j = \begin{cases} r_{j+1} & s_{j+1} \text{ 是最终状态} \\ r_{j+1} + \max_{a_{j+1}} Q(s_{j+1}, a_{j+1}, \theta^-) & s_{j+1} \text{ 不是最终状态} \end{cases}$ 
11:      以  $(y_j - Q(s_j, a_j, \theta))^2$  为损失函数训练  $Q$  网络
12:      每隔  $C$  步更新  $\theta^- \leftarrow \theta$ 
13:  end for
14: end for
输出:  $Q$  网络参数

```

引入经验缓存的目的在于降低序列关联度，同时提升数据运用效能^[43]。经验缓存涉及采集样本与采样样本两个环节，将搜集到的样本依时间顺序储存于特定结构中。如果经验缓存已满，训练模型会从原始数据中读取数据，出现训练不稳定或者过拟合的情况，为了解决这一问题智能体会实施缓存淘汰策略如先进先出，用新数据替代较早使用的数据提高训练模型的训练效率和稳定性。

2.2.2 基于策略梯度的深度强化学习算法

基于策略梯度的深度强化学习算法通过直接优化策略函数来学习如何在给定的状态空间中采取动作以最大化长期奖励。这种方法避免了传统基于值函数方法中需要学习的价值函数的复杂性和不稳定性，并且在某些场景下能够提供更好的性能。基于价值函数的深度强化学习的应用更适用于连续动作空间的环境而基于策略梯度的深度强化学习更适用于离散动作空间的环境。

在基于梯度函数的深度强化学习算法中，智能体通过与环境进行交互来收集经验数据。这些数据包括智能体所采取的动作、环境的状态以及所获得的奖励。然后，算法使用这些数据来计算策略的梯度，并根据梯度更新策略的参数^[44]。

具体来说，基于梯度函数的深度强化学习算法通常使用随机梯度下降来优化策略的参数。在每次更新参数时，算法都会计算策略梯度，并根据梯度的方向来

调整参数的值。通过不断地更新参数，算法可以逐渐改进智能体的策略，使其能够在环境中更好地执行任务。

其中DDPG算法是一种基于深度策略梯度的典型方法，DDPG算法基于Actor-Critic架构，该架构是一种强化学习中的常见架构，它由两个主要部分组成：演员（Actor）和评论家（Critic）。演员网络负责生成动作策略，它通过学习一个映射函数来预测在给定状态 s 下应该采取的动作 a ，采取的动作 a 会与环境进行交互生成新的状态值 s_{t+1} 以及奖励值 r_t 。评论家网络负责评估动作策略的好坏，它通过学习一个价值函数来预测在给定状态下采取某个最优动作 a_{t-best} 。在Actor-Critic架构中，演员和评论家网络是协同训练的，演员网络根据评论家网络提供的价值信息来改进动作策略，而评论家网络则根据演员网络的行为来更新价值函数。这种协同训练的方式可以提高学习效率和性能，在迭代过程中，演员网络负责策略网络参数 θ 的迭代更新并根据当前的状态生成动作概率分布，智能体根据这个分布选择一个动作执行，然后收集实际的奖励信号。批评家网络负责价值网络参数 ω 的迭代更新并根据智能体的行为和接收到的奖励来更新值函数 $Q(s_t, a_t)$ ，随后演员网络根据新的值函数评估来更新其策略。

算法 2-2 DDPG 算法^[45]

- 1: 初始化 Critic 网络 $Q(s, a; \omega)$, Actor 网络 $\mu(s; \theta)$
- 2: 初始化对应的目标网络 $Q' \leftarrow Q, \mu' \leftarrow \mu$
- 3: 始化经验缓存 R
- 4: *for* $i = 0 \rightarrow M$ *do*
- 5: 初始化起始状态 s_1
- 6: *for* $t = 0 \rightarrow T$ *do*
- 7: 根据当前策略 $a_t = \mu(s_t; \theta)$
- 8: 执行动作 a_t ，得到奖励值 r_{t+1} 和新的状态 s_{t+1}
- 9: 将 $\{s_j, a_t, r_t, s_{t+1}\}$ 存入 R 中
- 10: 从 R 中采样一批样本 $\{s_j, a_t, r_t, s_{t+1}\}$
- 11: $y_i = r_i + \gamma Q(s_{i+1}, \mu(s_{i+1}; \theta); \omega)$
- 12: 以 $\frac{1}{N} \sum_{i=1}^N (y_i - Q(s_i, a_i; \omega))^2$ 为损失函数训练 Critic 网络

- 13: 以 $\nabla_{\theta} J = \nabla_{\theta} \mu(s_t; \theta) \nabla_a Q(s_t, a_t; \omega)|_{a=\mu(s_t; \theta)}$ 为梯度更新 Actor 网络
- 14: 更新目标网络 $\omega' \leftarrow \rho\omega + (1-\rho)\omega'$ $\theta' \leftarrow \rho\theta + (1-\rho)\theta'$
- 15: end for
- 16: end for

2.3 源荷预测模型

源荷预测模型是一种重要的电力系统管理工具，在微电网能量管理中也起到重要的作用，可以用于预测未来一段时间内微电网系统的电源和负荷情况。随着新能源比例的逐渐提升，源荷预测模型的重要性日益增加，结合人工智能等先进技术预测效果会更加精准高效。源荷数据具有动态性、波动性、多维度和非线性等特点，源荷数据是实时变化的，受天气、季节、经济活动等多种因素的影响，例如在白天和深夜的电力需求量会有明显的差异，节假日的电力需求量也会低于工作日；对于可再生能源来说，每日的风力大小与光照强度都会影响发电量。

2.3.1 循环神经网络

微电网系统源荷预测任务难点在于数据的复杂性以及不确定性，传统的前馈神经网络信息是从输入层到输出层的单向传递，不适合直接用于时序数据的处理，为了解决这一问题，设计了具有短期记忆特性的循环随机网络，能够通过引入机制来控制信息的流动和记忆，从而有效地捕捉时序数据中的复杂关系。下面将对应用较多的循环神经网络进行详细的介绍：

(1) 典型循环神经网络

循环神经网络是一种基于序列数据的神经网络模型。相较于经典前馈网络，RNN 模型在处理序列数据时，可以捕捉到更丰富的语义和时序方面的信息^[46]。通常，序列数据是指具有有序内部元素的数据，如文字、文章和语句等。时间序列数据，例如一个月内的车流量、一个季度的天气情况等。

典型循环神经网络模型带有自反馈神经元，上一时刻的输出可以再次作为输入进行传输。其以 t 时刻输入样本 x_t 和 $t-1$ 时刻隐藏层状态 h_{t-1} 来更新 t 时刻隐藏层状态 h_t ，具体更新方式如式(2-11)、(2-12)所示。

$$h_t = f(W_x x_t + W_h h_{t-1} + b_n) \quad (2-11)$$

$$y_t = W_y h_t + b_y \quad (2-12)$$

式中 w_x , w_h , w_y 为各单元权重矩阵; b_n , b_y 为各单元偏置向量; $f(\bullet)$ 为 RNN 模型的激活函数。

典型循环神经网络模型的递归结构使得网络能够保持状态,即在时间步之间传递信息,这种能力使得 RNN 在捕捉时间序列数据中的时间依赖关系方面具有优势。但同时也存在一些局限性和不足之处,例如虽然 RNN 能够捕捉短期依赖关系,但在处理长期依赖关系时表现不佳,无法有效利用序列中的早期信息,对微电网的源荷预测效果产生较大影响;同时 RNN 在反向传播过程中会因为时间回溯导致梯度值非常小或非常大,破坏学习过程的稳定性。

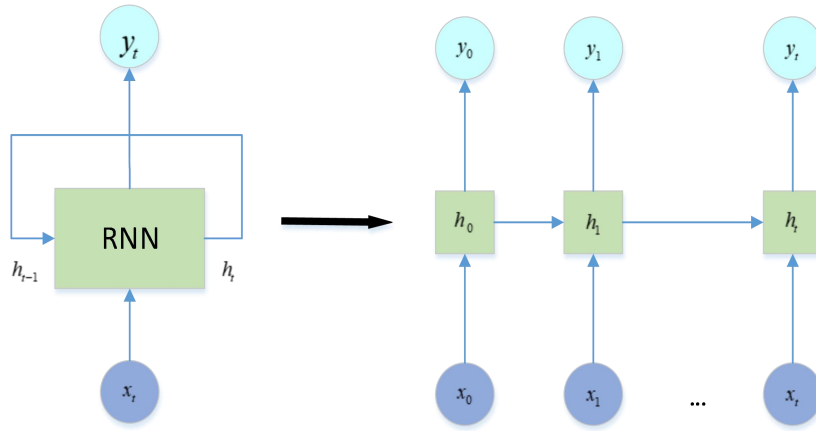


图 2-2 RNN 模型

(2)长短期记忆神经网络

LSTM 的诞生主要是为了应对传统 RNN 在处理长期依赖问题方面的局限问题。在 RNN 的基础上, LSTM 单元结构增添了更多复杂性,通过设立门控结构应对长期依赖问题,从而使其具备更好的短期记忆能力,相较于 RNN 表现更为优异^[47]。图 2-3 展示了 LSTM 模型的基本构成。

在 LSTM 单元中,保持门的作用是控制记忆模块的运作。输入门决定更新记忆单元的信息。输出门的主要职责是获取最近更新过的神经网络状况,并对记忆单元进行信息输出。LSTM 模型以 t 时刻的输入样本 x_t 、 $t-1$ 时刻的隐藏层状态 h_{t-1} 以及单元状态 c_{t-1} 作为输入,来更新 t 时刻三个门控 (f_t 、 i_t 、 o_t)、单元状态 c_t 以及隐藏层状态 h_t 。各单元结构更新方式如下所示:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2-13)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2-14)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (2-15)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2-16)$$

$$h_t = o_t \odot \tanh(c_t) \quad (2-17)$$

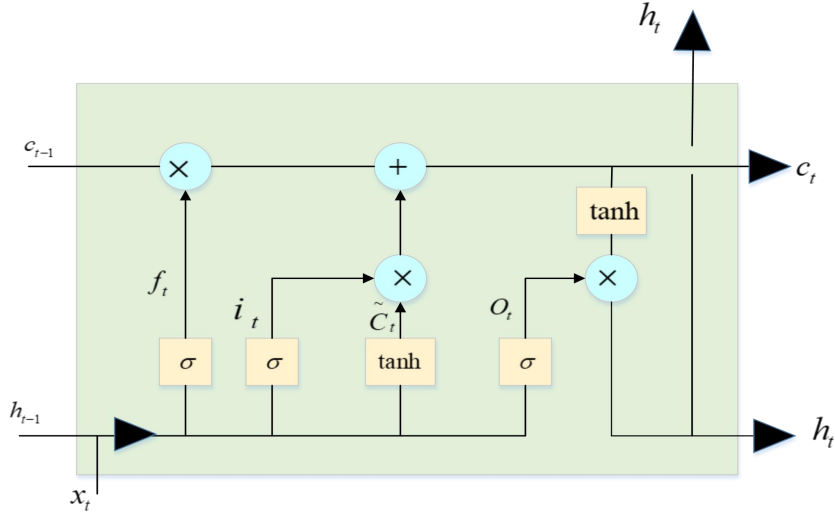


图 2-3 LSTM 模型

(3) 门控循环单元

在神经网络的演进中，针对 LSTM 结构，大部分研究文章都对其进行了适当调整，这些调整也可称为 LSTM 的衍生版本。较为显著的调整涉及门控循环单元（Gated Recurrent Units），简称 GRU，这是目前广泛应用的一种模型。在一定程度上，GRU 可视为对 LSTM 结构复杂性的优化。LSTM 具有三个独立的门结构而且参数较多，使得其训练过程更为复杂。GRU 这款模型仅有两个门控机制，当所有超参数均优化调整后，它们的性能表现相近。然而，GRU 架构更为简洁，适合训练样本较少的情况。

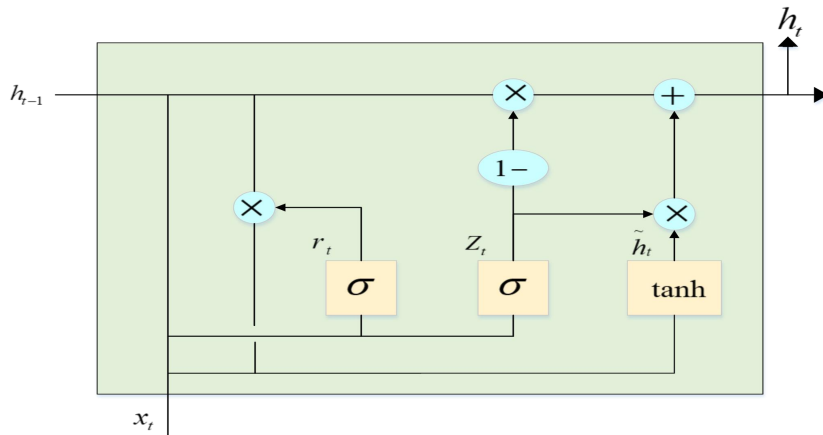


图 2-4 GRU 模型

GRU 在一定程度上简化了网络结构，去除了 LSTM 中的输出门和输入门，而是引入了一个更新门和一个重置门。更新门允许 GRU 网络决定在每个时间步

长应该如何整合新的输入信息，以及如何保留先前的记忆。重置门则有助于网络决定是否忽略先前的状态，并且有助于减少梯度消失问题。这两个门控制机制的引入使得 GRU 在某种程度上更容易训练，减少了梯度消失问题的影响^[48]。

此外，GRU 单元比 LSTM 单元更加简单，参数更少，因此在一些情况下可以更快地训练和运行。然而，由于 GRU 的设计更加简单，它可能在处理更复杂的长序列时，没有 LSTM 那么强大。因此在实际应用中选择使用 GRU 还是 LSTM 取决于具体任务的复杂程度和数据集的特征。GRU 的逻辑图如上图所示。GRU 模型以 t 时刻的输入样本 x_t 、 $t-1$ 时刻的隐藏层状态 h_{t-1} 作为输入，来更新 t 时刻的两个门控 (z_t, r_t) 以及隐藏层状态 h_t 。各单元结构具体更新方式如下所示：

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (2-18)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (2-19)$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \odot h_{t-1}, x_t] + b_h) \quad (2-20)$$

$$h_t = (1 - z_t) \odot \tilde{h}_t + z_t \odot h_{t-1} \quad (2-21)$$

2.4 本章小结

本章首先介绍了深度强化学习的背景以及马尔可夫决策过程。随后详细介绍了 DQN 算法和 DDPG 算法的理论基础与算法流程。接下来，详细阐述了几种循环神经网络，包括 RNN 模型、LSTM 模型以及 GRU 模型，通过对比分析了每个模型的优势以及应用场景。本文后续章节将采用神经网络和深度强化学习这两种关键技术来应对微电网能源优化管理问题。为此本章对相关技术进行了深入的梳理和提炼，为后续内容奠定了基础。

第3章 基于深度强化学习的微电网能量管理模型

3.1 微电网系统数学模型

本章构建包含光伏、风机、微型燃气轮机、储能设备以及负荷组件的并网型微电网，如图 3-1 所示。微电网中包括光伏发电设备、风力发电设备、微型燃气轮机、储能设备和负载。其中，光伏发电设备和风力发电设备属于可再生能源发电设备；微型燃气轮机是一种分布式发电设备；储能设备可以充电或放电。以上各种设备的状态可以由智能电表采集并发送到控制中心，控制中心使用设计的算法针对微电网的当前状态发出控制信号实现能源的优化管理。

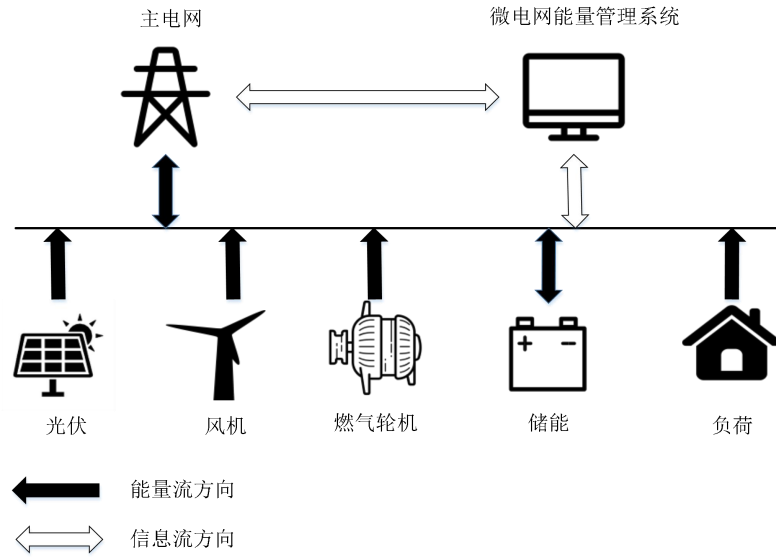


图 3-1 微电网框架

(1) 光伏发电模型

太阳能是一种清洁环保的能源，目前已经得到了广泛的使用。如今已在大范围内被广泛运用。把太阳能辐射转换成热、化学及电形式的储存与利用。光化学转换通过将太阳能辐射转化为氢的自由能等途径实现；太阳光能直接被太阳能板转换为电能，这个过程被称为光电转换；光伏发电中的光电转换现象最具代表性。光伏发电系统相较于传统电力系统，具备安全可靠、搭建便捷、环保等特点，适用于沙漠、平原、屋顶等场景。太阳能光伏发电的数学模型可以表示为^[49]：

$$P^{PV} = \eta SI (1 - 0.005(t_o - 25)) \quad (3-1)$$

$$P_{\min}^{PV} \leq P^{PV} \leq P_{\max}^{PV} \quad (3-2)$$

上式中， P^{PV} 为光伏发电功率（kW）； S 为光伏板面积（ m^2 ）； η 为转换效率

(%)； I 为辐射强度(kW/m^2)； t_0 为环境温度。

(2) 风力发电模型

风能是一种优秀的可再生能源，相较于光伏发电风力发电对选址及气候条件的挑剔程度更高，其随机性和不确定性较大。风力发电的数学模型可以表示为^[50]：

$$P^{WT} = \begin{cases} 0 & v < v_{in} \\ P_N (A + Bv + Cv^2) & v_{in} \leq v < v_N \\ P_N & v_N \leq v < v_{out} \\ 0 & v \geq v_{out} \end{cases} \quad (3-3)$$

$$A = \frac{1}{(v_{in} - v_N)^2} [v_{in}(v_{in} + v_N) - 4v_{in}v_N \left[\frac{v_{in} + v_N}{2v_N} \right]^3] \quad (3-4)$$

$$B = \frac{1}{(v_{in} - v_N)^2} [4(v_{in} + v_N) \left(\frac{v_{in} + v_N}{2v_N} \right)^3 - (3v_{in} + v_N)] \quad (3-5)$$

$$C = \frac{1}{(v_{in} - v_N)^2} [2 - 4 \left(\frac{v_{in} + v_N}{2v_N} \right)^3] \quad (3-6)$$

$$P_{min}^{WT} \leq P^{WT} \leq P_{max}^{WT} \quad (3-7)$$

上式中， P^{WT} 为风机的输出功率(kW)； P_N 为风力发电设备的额定功率(kW)； v 、 v_n 分别为实际风速和额定风速(m/s)； v_{in} 、 v_{out} 分别为切入风速和切出风速(m/s)， A 、 B 和 C 是风机参数。

(3) 微型燃气轮机模型

微型燃气轮机是一种小型的热力发动机，它利用气体流动的动力转换原理，将燃料的化学能转化为机械能，并通过发电机转化为电能。作为一种高效、清洁的能源转换设备，微型燃气轮机在分布式能源系统、混合能源系统、应急电源系统等领域具有广泛的应用前景。微型燃气轮机的数学模型可以表示为：

$$P^{MT} = F^{MT} H^{ng} \xi^{MT} \quad (3-8)$$

$$P_{min}^{MT} < P^{MT} < P_{max}^{MT} \quad (3-9)$$

上式中， P^{MT} 为微型燃气轮机的输出功率(kW)； F^{MT} 为燃气消耗量(Nm^3)； H^{ng} 为天然气热值(kWh/Nm^3)； ξ^{MT} 为发电效率(%)。

(4) 储能设备模型

储能设备是一种用于储存能量的装置，它能够将电能、热能、化学能等形式转化为便于存储和使用的能量，并在需要时将其转换回电能或其他形式的能量，目前主要包括电能、热能、化学能以及机械能四种储能设备。储能设备在微电网

中的应用非常广泛，能够实现能量的平衡管理、提高整体能源的利用效率和系统稳定性，随着储能技术的不断进步，储能设备在微电网中的应用将更加广泛，国内外已有众多 MW 等级的储能站点投入运营^[51]。其模型建立如下^[52]：

$$P_{\min}^{ESS} < P_t^{ESS} < P_{\max}^{ESS} \quad (3-10)$$

$$SOC_{t+1} = \begin{cases} SOC_t + \eta^{ch} P_t^{ESS} \Delta t / E^{ESS}, & P_t^{ESS} > 0 \\ SOC_t + P_t^{ESS} \Delta t / \eta^{dis} E^{ESS}, & P_t^{ESS} < 0 \end{cases} \quad (3-11)$$

$$SOC_{\min} \leq SOC_t \leq SOC_{\max} \quad (3-12)$$

上式中， P_t^{ESS} 为储能设备充放电功率； $P_{\max}^{ESS}$ 、 $P_{\min}^{ESS}$ 分别为储能设备功率的上下限(kW)；SOC 为储能设备的荷电状态； $SOC_{\max}$ 、 $SOC_{\min}$ 分别为储能设备荷电状态的上下限； Δt 为充放电时间(h)； η^{ch} 、 η^{dis} 分别为充放电系数(%)； E^{ESS} 为储能设备容量(kWh)。

(5) 主电网模型

微电网系统须与主电网协同运作，实现功率互换，分布式能源的参与可以降低运行成本，由于分布式能源具有较大的不确定性为了确保实际情况下的用电稳定，还需要主电网进行供电。主电网的功率约束为：

$$P_{\min}^{MG} \leq P^{MG} \leq P_{\max}^{MG} \quad (3-13)$$

上式中， P^{MG} 代表微电网与主电网交易的功率(kW)； $P_{\max}^{MG}$ 、 $P_{\min}^{MG}$ 分别为微电网与主电网交易的功率的上下限(kW)。

3.2 基于深度强化学习的微电网能量管理模型

本节将会在深度强化学习的框架下根据微电网系统的数学模型建立基于深度强化学习的微电网能量管理模型，并基于该模型使用深度强化学习算法来实现微电网能量管理，从而在算例分析中验证深度强化学习方法的优势。

在第二章中介绍了为了解决实际应用问题需要将问题建模转化为一个马尔可夫决策过程。智能体首先需要对微电网的状态 $s_t \in S$ 进行观察并根据动作选择策略 π ：选择一个动作 $a_t \in A$ 用于调节微电网中设备的运行状态以及与主电网交

易的功率值，随后转移的下一状态 $s_{t+1} \in S$ ，并获得奖励 r_t 。

(1)状态空间

深度强化学习算法可以实现能量调度的依据是智能体可以接收的外界环境的信息以及变化，根据环境变化的不同生成不同的决策，所以首先要定义智能体的状态空间^[53]。微电网能量管理状态空间可定义为：

$$S_t = [P_t^{PV}, P_t^{WT}, P_t^L, c_t, SOC_t] \quad (3-14)$$

式中， P_t^L 为在时刻 t 的负荷需求功率， c_t 为在时刻 t 的实时电价， SOC_t 为在时刻 t 的储能荷电状态。

(2)动作空间

动作空间的概念需基于智能体所能实施且需达成的动作来设定，以达到优化目标的目的。因此，降低运行成本、减少污染排放、事先功率平衡等优化目标就可以通过微电网能量管理系统控制可控设备实现。微电网能量管理的动作空间定义为：

$$A_t = [P_t^{MG}, P_t^{ESS}, P_t^{MT}] \quad (3-15)$$

式中 P_t^{ESS} 为在时刻 t 的储能充放电功率

(3)奖励函数

深度强化学习智能体可以通过执行决策获得奖励信号，为了获得更多奖励会不断更新决策。奖励信号充当优化目标与深度强化学习算法之间的纽带，其生成规则被称为奖励函数^[54]。设定奖励函数并非唯一固定，各种奖励函数对算法实际效能、收敛速度及性能表现皆存在程度不一的影响。需要根据不同环境和优化目标进行设计。

(4)状态转移概率

状态转移概率表示在当前环境状态下，经过行为干预后，系统可能跃迁至其他状态的概率分布规律^[55]。在微电网环境中，光伏发电、风力发电、用户负荷需求、电价以及储能电荷状态这些因素均呈现不断变动的特征。微电网的状态转移概率表达式为：

$$P[S_{t+1} | S_t, A_t] = P[P_{t+1}^{PV} | P_t^{PV}] \cdot P[P_{t+1}^{WT} | P_t^{WT}] \cdot P[P_{t+1}^L | P_t^L] \cdot P[c_{t+1} | c_t] \cdot P[SOC_{t+1} | SOC_t, P_t^{ESS}] \quad (3-16)$$

式中， P_{t+1}^L 为在 $t+1$ 时刻的负荷需求功率， SOC_{t+1} 为在时刻 $t+1$ 的储能荷电

状态。由于在微电网中分布式能源的不确定性以及实时电价的波动性导致微电网环境下的状态转移概率也具有很强未知性，为了解决这一问题，需要智能体通过大量的训练，根据训练做出相应决策并不断完善策略。

3.3 算例分析

为了验证深度强化学习方法实现微电网能量管理的性能，体现深度强化学习方法在适用于不确定场景的优势，本节在一个家庭微电网测试场景中建立了深度强化学习能量管理模型，并使用 2.2 节介绍的深度强化学习算法：DDPG 和 DQN 对智能体进行训练，以实现能量优化管理。

3.3.1 仿真实验数据及参数定义

在这次仿真实验中建立了一个微电网模型，该微电网模型包含用户负荷以及储能设备，储能设备的参数如表 3-1 所示。本节所使用的负荷数据来自于开源数据平台 Open Power System Data。实验在 Python3.6 和 TensorFlow 框架下编写程序，进行仿真实验。仿真实验所用的设备处理器为 AMD Ryzen7 5800H,显卡为 NVIDIA GeForce RTX3050。

表 3-1 储能设备参数

参数	数值	参数	数值
$P_{\min}^{ESS}$ (kW)	0	$P_{\max}^{ESS}$ (kW)	30
$SOC_{\min}$ (%)	20	$SOC_{\max}$ (%)	90
$\eta_{\max}^{ch}$ (%)	95	η_{MAX}^{dis} (%)	95

该仿真场景的状态空间定义如式(3-17)所示，动作空间如式(3-18)所示，状态转移概率如式(3-19)所示，奖励函数以运行成本最小化为目标进行设计，如式(3-20)所示，运行成本根据(3-21)计算。

$$S_t = [P_t^L, c_t, SOC_t] \quad (3-17)$$

$$A_t = [P_t^{ESS}] \quad (3-18)$$

$$p[S_{t+1} | S_t, A_t] = p[P_{t+1}^L | P_t^L] \cdot p[SOC_{t+1} | SOC_t, P_t^E] \quad (3-19)$$

$$R(S_t, A_t) = -COST_t \quad (3-20)$$

$$COST_t = \sum_t^T P_t^G \cdot c_t \cdot \Delta t \quad (3-21)$$

3.3.2 仿真结果与分析

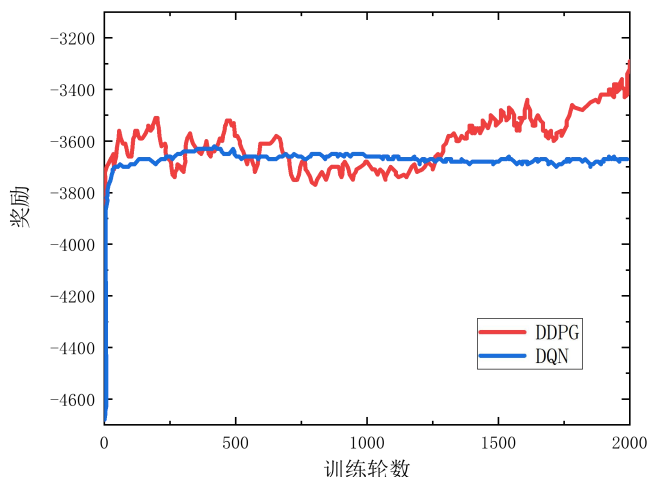


图 3-2 DDPG 算法和 DQN 算法训练效果对比

图 3-2 为 DDPG 和 DQN 两种算法的训练过程对比图，总体上来看两种算法在经过 2000 轮的训练过程后奖励曲线均呈现上升的趋势，说明深度学习算法在训练过程中可以完成降低微电网使用成本的进而完成能量管理的任务。分开来看，DDPG 算法在上升过程中波动较大，约在 1750 集后奖励逐渐收敛在-3400 左右其仍呈现一定的上升趋势。而 DQN 算法在训练的初始阶段奖励上升速度很快，在 500 集后奖励保持相对稳定最终收敛在-3700 左右，但在奖励值并非最优值。通过对比可以做出以下结论：DQN 算法能够较快使智能体通过训练获得奖励，但是使用的策略也相对固定获得的奖励也相对固定。而 DDPG 算法对动作策略的探索则更加积极，需要更多的时间进行探索，但同时也会获更多的奖励。

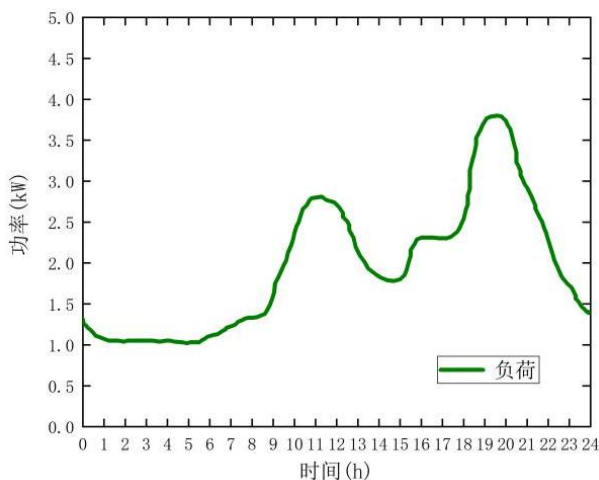


图 3-3 负荷需求

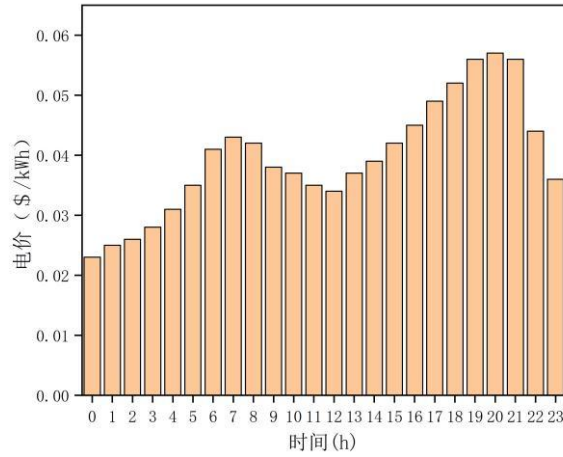


图 3-4 实时电价

为了确认经过训练的智能体在各种环境下能否作出正确的决策,本次仿真实验计划随机选择一个典型日进行模拟。需要注意的是,典型日的数据与用于训练智能体的数据集并不相同。图 3-3、3-4 分别为典型日的住宅负荷需求与实时电价。

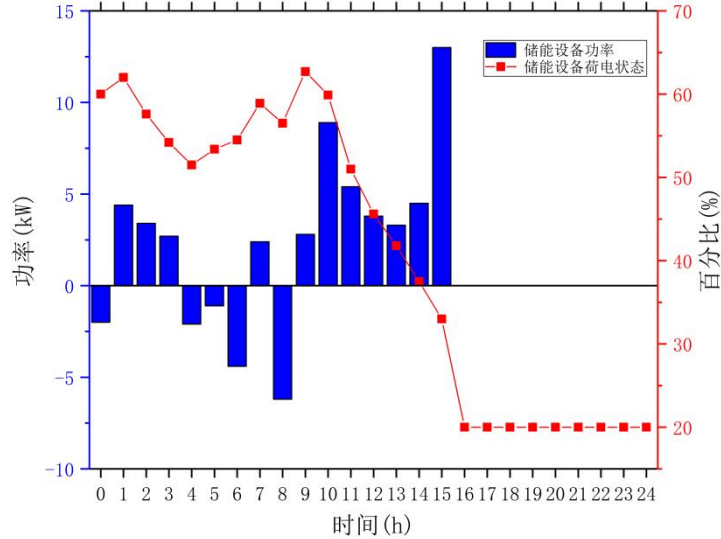


图 3-5 DDPG 算法调度结果

由图3-5可以看出DDPG智能体的能量管理策略是结合了用户负荷以及实时电价。首先在4时至7时电价较低且负荷也较低的情况下,储能装置进行充电。随后在9时至16时,随着用户负载的提升储能开始进行放电,降低了从主电网的购电量。

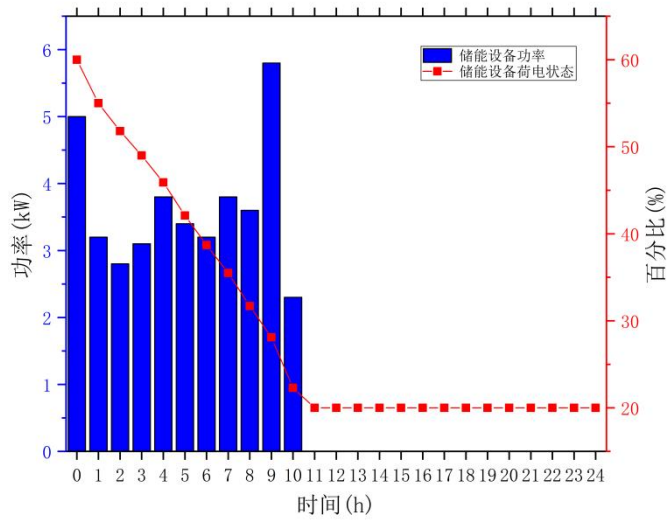


图 3-6 DQN 算法调度结果

图 3-6 为 DQN 算法调度结果，与 DDPG 智能体相比，DQN 智能体在起始阶段 0-10 时，就利用储能装置为负载持续提供电力。通过减少对主电网的购电完成降低成本的目标，但在后续 10 时至 20 时的用电高峰及电价高峰时，储能装置的荷电状态约在 20%左右处在过放状态，在这种情况下储能装置无法供能，只能使用电网供能。

通过两种智能体的对比可以发现，DQN 智能体的能量管理策略目光较为短浅局限于更早的获得奖励，而 DDPG 智能体的眼光更为长远，能够更合理的对能量进行管理。

表 3-2 中展示了两种方法在典型日的优化效果。DDPG 智能体的运行成本为 1.69 美元，DQN 智能体的运行成本为 2.27 美元。表明 DDPG 算法在训练过程中通过不断的探索，实现了更为优秀的能量管理策略。实验也表明了深度强化学习算法通过学习目标对象的相关数据集后可以完成对不确定环境下的能量管理。

表 3-2 两种方法的优化效果对比

	DDPG	DQN
运行成本（\$）	1.69	2.27

3.4 本章小结

本章针对基于深度强化学习的微电网能量管理策略进行研究，首先建立了微电网的数学模型，并在深度强化学习框架下建立了微电网能量管理模型，使用深

度强化学习算法实现了微电网能量管理，最后利用仿真实验进行了算例分析。在算例分析中，DDPG 算法和 DQN 算法相比成本降低了 25.56%，证明了 DDPG 算法在减少运行成本方面更具有优势，同时仿真结果表明与 DDPG 算法相比 DQN 算法能够更积极的探索优化策略，在能量调度方面更为合理。

第4章 基于 LRODDPG 的单微电网能量管理

4.1 引言

在上一章中通过对一个家庭微电网模型的仿真分析，证明了利用深度强化学习算法可以完成微电网的能量管理，但该模型没有考虑风电、光伏等可再生能源，状态转移概率主要以用户负荷概况分布，优化目标较为简单，智能体的训练难度也较低，对奖励函数的设置要求不高。在更为复杂的环境状态下，智能体如果仅仅将运行成本作为奖励信号返回给智能体，可能会出现风电光伏等可再生能源匮乏的情况下，微电网在储能装置高功率运行的情况下依然不能满足实际需要，仍然需要从主电网购电确保用电可靠性，此时因为运行成本过高智能体会给出较低的奖励；同样的在可再生能源充沛的情况下，智能体会因为运行成本较低无论使用什么策略均给出较高的奖励。

深度强化学习算法需要依靠奖励函数的反馈不断学习调整策略，针对不同的环境需要适当调整奖励函数的设置，这样才能保证智能体的训练效果以及算法的收敛效果。在本节中将对第三章中效果更好的 DDPG 算法进行优化，在模型加入长短期记忆神经网络提取源荷数据并对奖励函数进行重新设计，进一步完善智能体的优化效果。

4.2 单微电网能量管理模型

4.2.1 优化目标与奖励函数定义

在本章的研究中，优化目标包括降低微电网运行成本以及减少碳排放量。首先微电网的运行成本 $COST_t$ 为：

$$COST_t = \sum_t^T P_t^G \cdot c_t \cdot \Delta t \quad (4-1)$$

c_t 为在 t 时刻的实时电价，电网交换功率 P_t^G 根据式(4-2)的平衡功率约束确定：

$$P_t^G = P_t^{PV} + P_t^{WT} + P_t^{ESS} - P_t^L \quad (4-2)$$

为了直观体现碳排放量的优化目标，将微电网中使用光伏、风电以及储能电池供电的情况转化为碳排放量的优化量，在微电网中分布式能源的出力不会产生碳排放，当分布式能源充沛时会减少从主电网的购电量，从而达到减少碳排放

的目标。微电网中分布式能源的供能与用户负载功率的比例定义为自平衡率 R_t^{sf} ：

$$R_t^{sf} = \frac{P_t^d}{P_t^{lb}} \quad (4-3)$$

$$C_t = \tau_e \cdot P_t^d \cdot \Delta t \quad (4-4)$$

在上式中 P_t^d 为分布式能源在 t 时刻的供能(kW)， P_t^{lb} 为储能电池的充电功率和用户的负载需求， C_t 为 t 时刻的碳排放减少量， τ_e 为碳排放转换系数^[56]。当分布式能源提供的能量越接近于负载需求和电池充电功率时，说明在微电网中清洁能源的占比越高，碳排放减少量也就越多。

基于本章的优化目标，奖励函数可以简要定义为：

$$R(S_t, A_t) = f(R_t^{sf}, COST_t) \quad (4-5)$$

式中 $f(\cdot)$ 表示根据自平衡率和运行成本作为因变量的奖励函数。

4.2.2 状态与动作空间定义

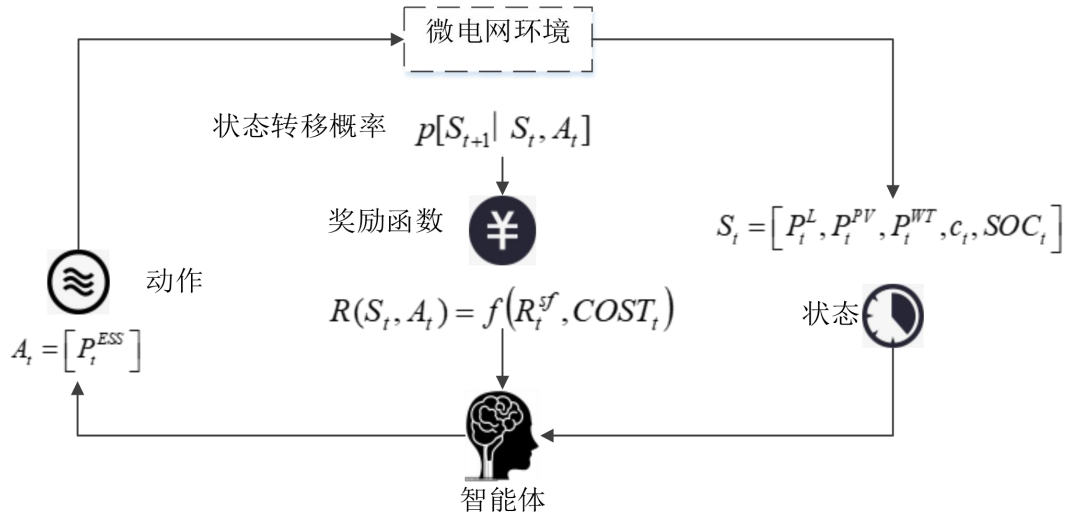


图 4-1 微电网深度强化学习能量管理模型

微电网环境中提供给智能体的环境信息包括负载功率、光伏风力发电功率、实时电价以及储能设备荷电状态，根据优化目标和环境信息，微电网能量管理模型状态空间可以定义为：

$$S_t = [P_t^L, P_t^{PV}, P_t^{WT}, c_t, SOC_t] \quad (4-6)$$

微电网中的可控设备为储能装置，在 t 时刻储能装置的充放电功率即为 t 时刻的动作空间。微电网能量管理模型的动作空间可以定义为：

$$A_t = [P_t^{ESS}] \quad (4-7)$$

根据状态空间和动作空间可得知，状态转移概率定义为：

$$p[S_{t+1}|S_t, A_t] = p[P_{t+1}^{PV}|P_t^{PV}] \cdot p[P_{t+1}^{WT}|P_t^{WT}] \cdot p[P_{t+1}^L|P_t^L] \cdot p[c_{t+1}|c_t] \cdot p[SOC_{t+1}|SOC_t, P_t^{ESS}] \quad (4-8)$$

4.3 基于 LRODDPG 的单微电网能量管理方案

4.3.1 基于 LSTM 的源荷时序数据提取

深度强化学习智能体训练的难点在于风电、光伏以及用户负荷的不确定性较强，微电网的外部环境变化影响智能体状态空间，对智能体的输出动作也会产生较大影响，加入神经网络可以通过训练对随机性较大的数据进行预测，可以有效的解决分布式能源及负荷不确定的问题。

在神经网络模型的选择上对比了 RNN、LSTM 和 GRU 三种循环神经网络在同一数据集下训练的负荷预测结果，为了评估三种模型的负荷预测性能，本文采用平均绝对百分比误差(MAPE)、均方根误差(RMSE)、平均绝对误差(MAE)这三种指标评估模型预测效果。其中 MAPE 能够反映出预测值相对于实际值的比例误差，更能体现预测值与实际值之间的关系；RMSE 能够反映出预测值与实际值之间的偏差程度；MAE 能够反映出预测值与实际值之间的绝对偏差。上述三种指标越低代表模型预测效果越好，具体计算公式如下：

$$MAPE = \frac{100\%}{m} \sqrt{\sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right|} \quad (4-9)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (4-10)$$

$$MAE = \frac{1}{m} \sqrt{\sum_{i=1}^m |y_i - \hat{y}_i|} \quad (4-11)$$

式中，m 为样本总量； y_i 、 $\hat{y}_i$ 分别为第 i 个时间点负荷的真实值和估计值。

时序模型的参数指标和超参数设置如表 4-1、4-2 所示。

表 4-1 时序模型参数指标设置

参数指标	参数设置
输入步长	12
输出步长	6
损失函数	MSE
优化器	Adam

续表 4-1 时序模型参数指标设置

参数指标	参数设置
迭代次数	200
评价指标	MAPE\RMSE\MAE

表 4-2 时序模型超参数设置

时序模型	层一节点数	层二节点数	批量大小	丢弃率	学习率
RNN	160	160	128	0.100	0.010
GRU	160	160	128	0.100	0.010
LSTM	160	160	128	0.100	0.010

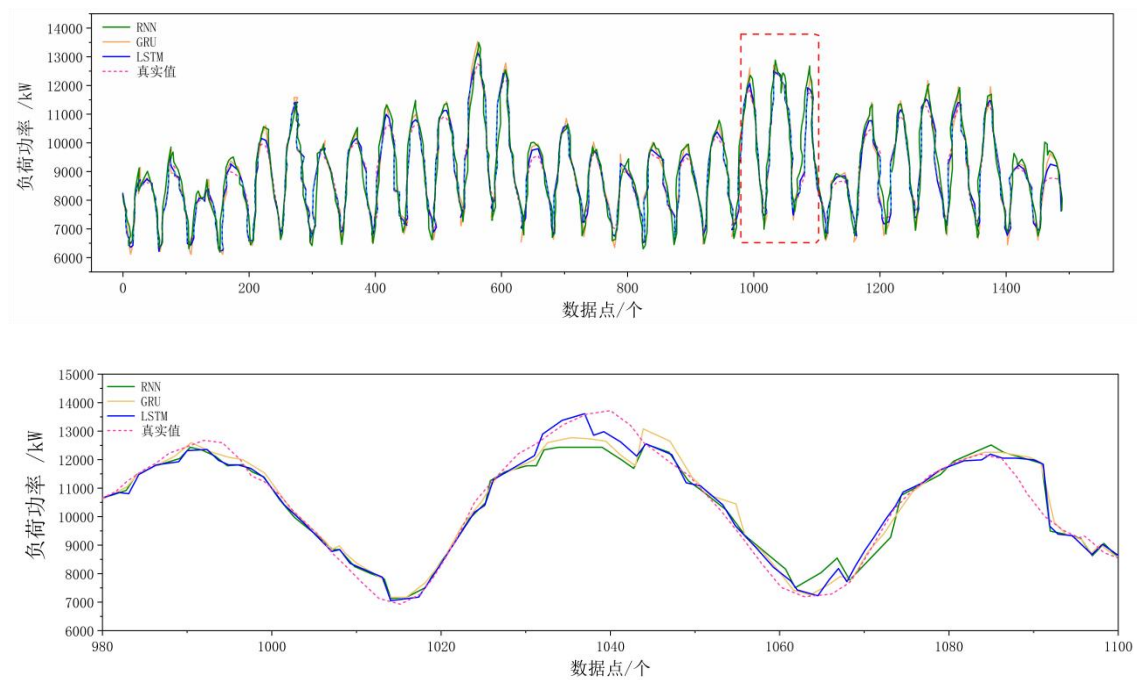


图 4-2 模型预测结果对比(上)和局部放大对比(下)

图 4-2 展示三种时序模型的预测效果对比,通过对比可以发现,LSTM 模型的预测曲线和真实值的变化趋势最接近,预测曲线在波峰波谷区域与真实负荷曲线更贴合。

表 4-3 展示了三种时序模型的预测结果误差。通过对比数据可以发现,LSTM 预测模型的 MAPE、RMSE、MAE 三项结果均为最低,LSTM 模型相较于 RNN 和 GRU 模型 MAPE 降低了 26.56%和 6.76%,RMSE 降低了 22.75%和 4.59%,MAE 降低了 27.22%和 5.29%。综上所述,LSTM 模型的预测效果最好,本章将使用 LSTM 模型进行对源荷数据的预测。

表 4-3 模型预测结果误差

模型	MAPE(%)	RMSE	MAE
RNN	1.653	197.257	141.924
GRU	1.214	152.374	103.288
LSTM	1.302	159.714	109.056

在时序预测中，可以利用 LSTM 可以学习长期依赖信息的特点，只保留有用的信息进行预测而不是全部记忆^[57]。在本文中训练集的时间跨度较长，光伏、风力和负荷的关联性较强使用 LSTM 神经网络对这些数据进行提取并进行预测，将预测的数据和智能体的状态空间共同输入给 DDPG 算法，通过算法训练做出决策完成微电网的能量调度。结构图如图 4-3 所示。

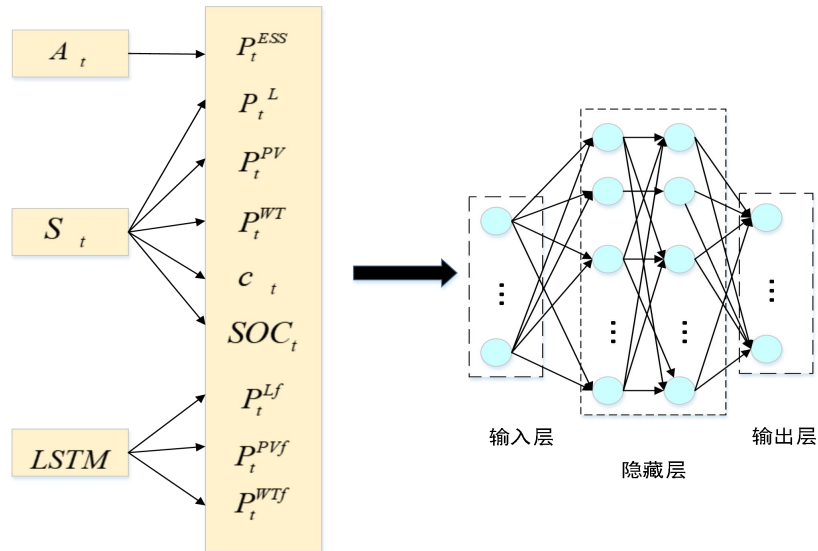


图 4-3 DDPG 网络输入结构图

图中 P_t^{Lf} ， P_t^{PVf} ， P_t^{WTf} 分别为 LSTM 神经网络预测的负荷功率、光伏发电功率和风力发电功率。

4.3.2 奖励函数优化设计

智能体通过接收外界环境状态以及获得的奖励输出动作完成一次训练并更新状态和奖励，不断循环往复最终实现能量调度。由此可以看出算法的质量取决于状态、奖励和动作三个方面，本节以奖励函数为切入点，通过对奖励函数进行优化，以提升深度强化学习算法的性能，从而更好地助力智能体的训练进程。

在奖励函数的设置上，常规算法只设置一个奖励函数反馈奖励信号给智能体，

本文的优化奖励函数将设置即时奖励和完成奖励两个不同的奖励函数对智能体反馈奖励信号。即时奖励只对智能体的当前动作进行评估，而完成奖励则在每一个训练集完成训练后对智能体完成的状态进行总体评价，两个奖励函数在不同时刻以及不同环境反馈不同的奖励，对智能体的训练效率及算法的收敛速度都有提高。下面就详细介绍两种奖励的设计。

表 4-4 奖励优化强化学习方法

奖励优化深度强化学习方法

```

1: 设置惩罚系数  $L_1$  和奖励系数  $L_2$ 
2: for 每个训练集 do
3:   初始化即时奖励  $r_t^{im}$  和完成奖励  $r_t^f$ 
4:   观察状态  $S_t$  和动作  $A_t$ 
5:   计算  $t$  时刻的即时奖励:  $r_t = r_t^{im}$ 
6:   计算  $t$  时刻的完成奖励:  $r_f = r_f + (r_1 + r_2 + \dots + r_n)$ 
7:   if 训练集结束  $t=T$ 
8:      $r_t = r_t^f$ 
9:   else
10:     $r_t = r_t^{im}$ 
11:   end if
12: end for

```

即时奖励的设置较为常规，按照前文的描述奖励的优化目标为降低运行成本和提高系统的自平衡率，奖励则为这两个的指标乘以惩罚系数，惩罚系数在深度强化学习中通常指算法中的折扣因子(discount factor)，该系数决定了对奖励的折扣程度^[58]。智能体的决策过程会受到折扣因子的影响，较小的折扣因子会让智能体对当前奖励更为重视，反之较大的折扣因子会让智能体更关注未来的奖励。综上所述即时奖励的公式为：

$$r_t^{im} = \left[\left(\left(\sum_{i=1}^t \frac{COST_t}{COST_t^N} \right) / t \right) + \left(\sum_{i=1}^t R_i^{sf} \right) / t \right] * L_1 \quad (4-12)$$

式中 $COST_t^N$ 为去除光伏风力等可再生能源，用户负荷只通过主电网购电的成本， $\left(\sum_{i=1}^t \frac{COST_t}{COST_t^N} \right) / t$ 用实际成本除以单一购电成本，成本节约度与比值成正比，

比值越高节约程度越高。 $\left(\sum_{i=1}^t R_i^{sf} \right) / t$ 表示 1 到 t 时的自平衡率。 L_1 为惩罚系数，

通过对数据集大小及模型复杂程度等进行综合考虑，经过调整最终确定惩罚系数

值为-15。

完成奖励的设置需要通过整体评估智能体的输出动作和当前环境的匹配程度，在一个环境状态中智能体做出的动作策略越好获得的奖励值也就越高。在完成奖励的奖励函数设置上，本文通过采用自平衡率的高低作为评估智能体动作优劣的标准。计算的方式如下：

$$r_t^f = \begin{cases} \left(1.2 \cdot \frac{P_t^{ESS-dis}}{P_t^{lb}}\right) \cdot L_2 & \text{if } R_t^{sf} \geq 90\% \\ \left(\frac{P_t^{ESS-dis}}{P_t^{lb}}\right) \cdot L_2 & \text{if } R_t^{sf} < 90\% \end{cases} \quad (4-13)$$

式中当微电网自平衡功率大于百分之九十时， $\frac{P_t^{ESS-dis}}{P_t^{lb}}$ 会乘以一个系数获得额外奖励其中 $P_t^{ESS-dis}$ 为储能电池的放电容量，而当自平衡功率小于百分之九十时不会获得额外奖励但依旧会获得奖励，通过奖励函数的设定可以鼓励智能体通过提高自平衡率以获得更高奖励。

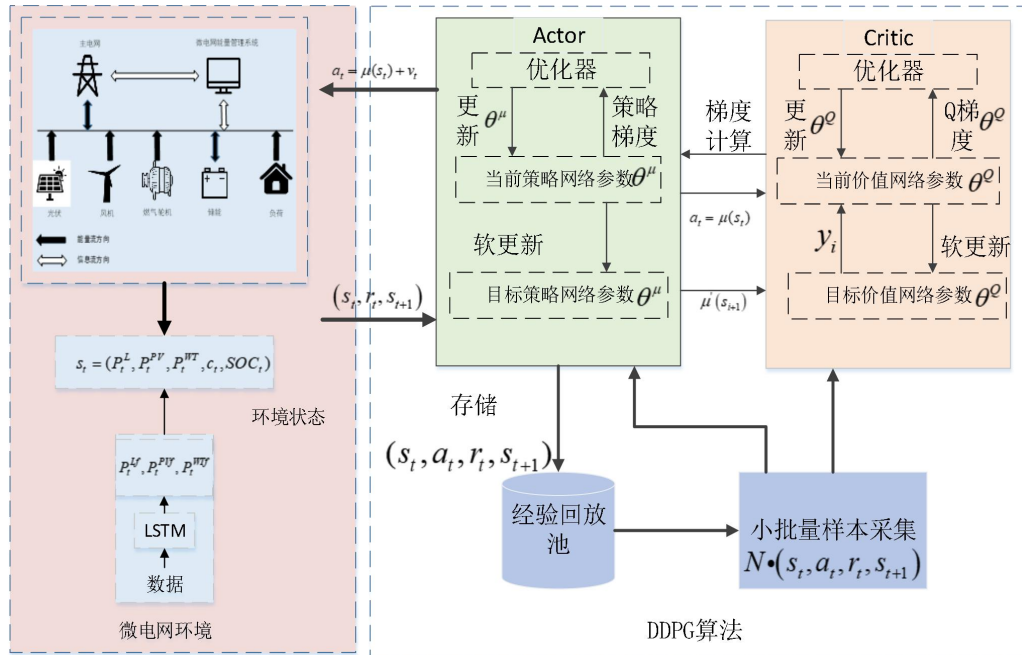


图 4-4 LRODDPG 算法的训练过程框图

图 4-4 为 LRODDPG 算法的训练过程框图，如图所示首先利用 LSTM 模型预测了未来调度时段的光伏、风力和负荷，构成状态空间。随后通过 DDPG 算法输出调度动作作用于微电网环境中，并接收到反馈奖励和下一阶段的环境状态，

实现算法和微电网环境的交互。

4.4 实验结果分析

4.4.1 仿真实验数据

在本节中使用上文所提出的 LRODDPG 算法以及常规 DDPG 算法对同一个微电网系统进行基于深度强化学习的微电网能量调度,通过对算法收敛性以及实际调度效果的对比验证 LRODDPG 算法的提升。本节的实际仿真场景包含风力、光伏、储能设备以及负载,相较于 3.3 节的微电网仿真模型,增加了风力光伏等不可控功率设备,微电网环境更为复杂同时也更符合更适用于实际场景。实际参数设置如表 4-5 所示:

表 4-5 微电网参数设置

设备类型	参数	数值	参数	数值
风力发电	$P_{\min}^{WT}(\text{kW})$	0	$P_{\max}^{WT}(\text{kW})$	500
光伏发电	$P_{\min}^{PV}(\text{kW})$	0	$P_{\max}^{PV}(\text{kW})$	600
负荷	$P_{\min}^L(\text{kW})$	0	$P_{\max}^L(\text{kW})$	1200
储能设备	$SOC_{\min}$	0.2	$SOC_{\max}$	0.9
	η^{ch}	0.95	η^{dis}	0.95

强化学习过程是一个不断探索的过程,需要通过大量训练集的训练网络参数后才能用于实际决策,本章所使用的风力光伏以及负荷数据来自于开源数据平台 Open Power System Data,仿真实验所用的设备处理器为 AMD Ryzen7 5800H,显卡为 NVIDIA GeForce RTX3050,在 Python 3.6 和 TensorFlow 框架下编写程序,进行仿真实验。通过不断实验最终确定智能体使用两层隐藏层,每层有 64 个神经元,中间层采用 RELU 激活函数,Actor 网络使用 softmax 激活函数输出每个动作的概率分布。LSTM 神经网络的参数设置为:学习率设置为 0.005,由三个隐藏层组成,每个隐藏层的神经元个数为 128,使用 Adam 优化器。

4.4.2 训练效果对比

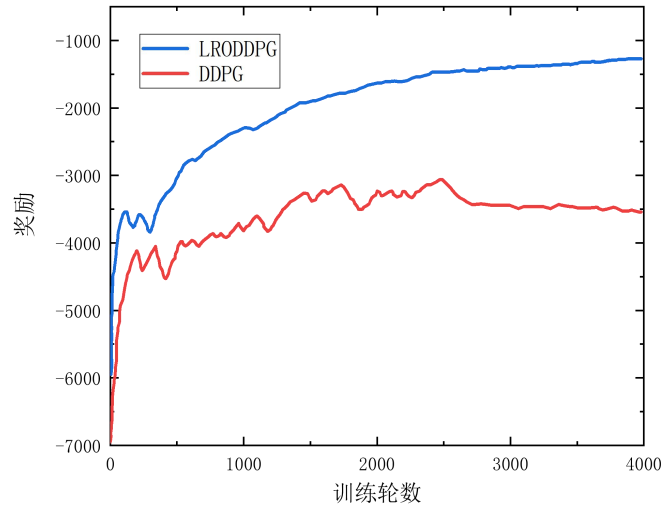


图 4-5 两种算法的训练效果对比

图 4-5 为是使用经过优化的 LRODDPG 算法和常规的 DDPG 算法经过 4000 轮学习后所获得的奖励，两种算法曲线均呈现波动上升的趋势，说明深度强化学习算法可以通过训练不断获得更高的奖励。但常规 DDPG 算法的波动性较大约在 2600 集左右奖励开始收敛于-3500 左右，且该奖励并非智能体获得的最高奖励属于次优解，在后期呈现一定的下降趋势；而 LRODDPG 算法在训练初期具有一定波动性，随后呈现稳定的上升趋势，约在 2000 集后奖励逐渐开始收敛于-1500 左右，且通过训练的智能体随着训练轮数的提高，通过不断地探索奖励值在不断提高最终可以获得最优解。从奖励函数角度分析 LRODDPG 算法可以获得更高的奖励值；从迭代次数分析 LRODDG 算法可以更快的完成收敛，说明了优化算法学习效率更高从而减少训练的时间；从收敛后的稳定性角度分析，除了初期两种算法均呈现一定波动性外，LRODDPG 算法的稳定性远高于常规 DDPG 算法。

如图 4-6 所示将 LRODDPG 和 RODDPG 算法的训练过程进行单独比较，可以发现两种算法的收敛趋势大致相同，但加入 LSTM 神经网络的 LRODDPG 算法收敛后的效果更为平滑，且获得的奖励更高，通过对比验证说明了引入 LSTM 对微电网外部的随机环境进行预测，可以有效的提升算法的稳定性，对训练过程起到增益作用。

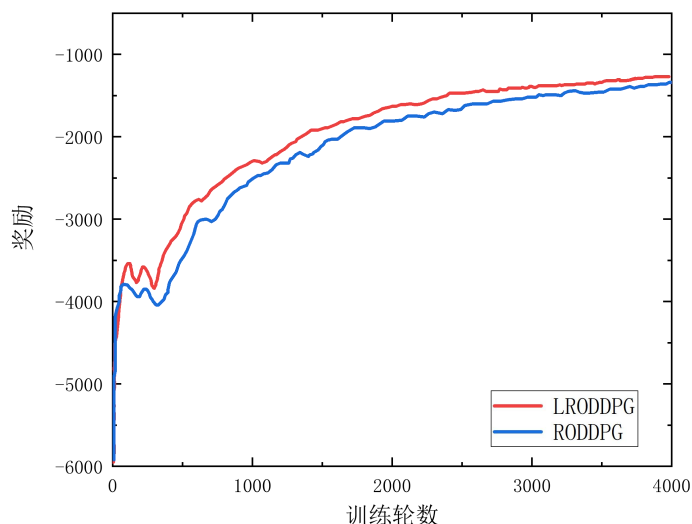


图 4-6 LRODDPG 与 RODDPG 训练效果对比

4.4.3 调度结果对比

仅仅通过算法收敛获得奖励的方法不能直观的展现出算法的实际性能, 所以将训练好的智能体选取随机一个微电网典型日进行决策。图 4-7、4-8、4-9 分别为光伏、风力的发电出力和负荷的实际情况以及基于 LSTM 神经网络的预测值, 由图可以分析得到光伏发电主要集中在 7-17 时而风力发电由于不受光照影响可以实现全天出力。将风电光伏设备的出力情况和用户负荷对比, 计算功率缺口后可以发现, 可再生能源的功率可以基本满足负荷需求, 减少了从主电网的购电量。

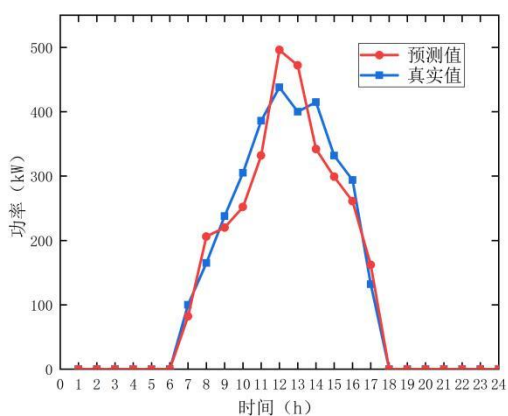


图 4-7 光伏出力情况及预测曲线

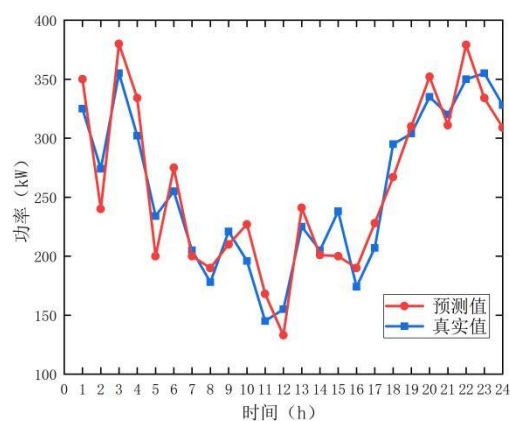


图 4-8 风力出力情况及预测曲线

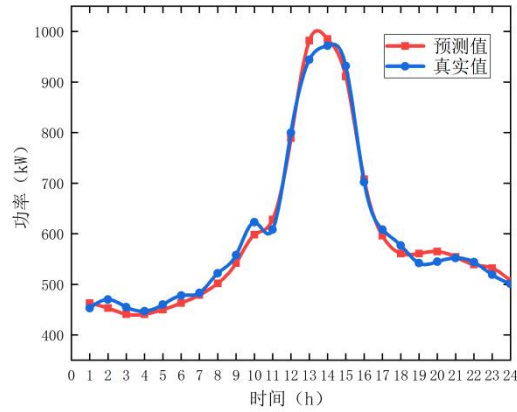


图 4-9 负荷需求及预测曲线

在训练 LSTM 模型的过程中，光伏、风力和负荷的在训练集上的网络损失均方差(MSE)分别为 0.00244、0.00363 和 0.00128，其中 MSE 是一种常用的回归问题评估指标，用于衡量模型预测值与实际值之间的差异 MSE 的值越小，表示模型的预测值与实际值越接近，模型的性能越好。使用通过训练后的 LSTM 模型进行预测后，光伏、风力、负荷的平均绝对百分比误差(MAPE)分别为 21.6%、27.3%和 5.4%，可以用于模型预测。

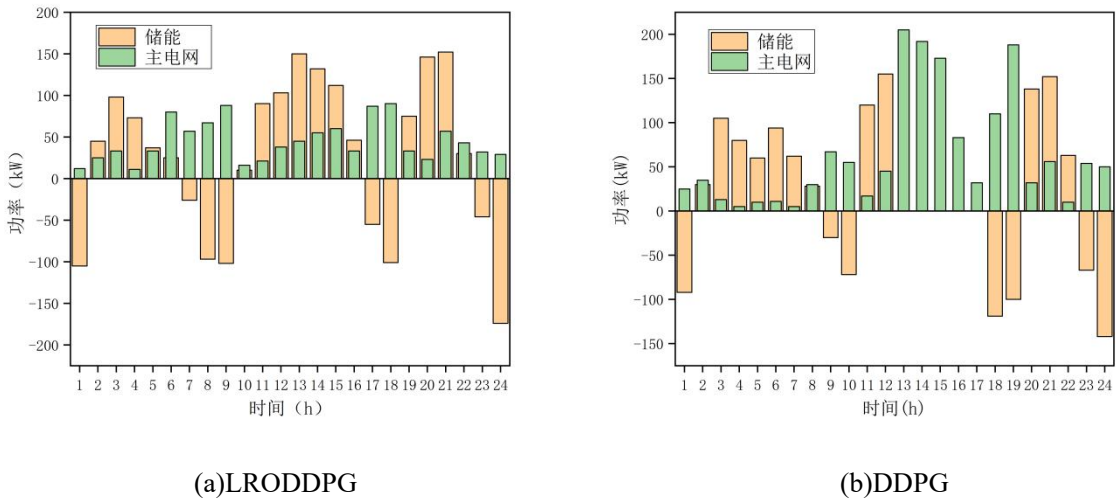


图 4-10 两种算法的调度结果

图 4-10 为两种方法在一个微电网典型日的调度结果，通过对比可以发现 LRODDPG 智能体可以在可再生能源充沛时对储能装置进行充电并且在用户负荷高时参与调度，很大程度的减少了从主电网购电的需求，降低的使用成本同时也降低的自平衡率，更好的完成了优化调度目标，而常规 DDPG 智能体在电价较低的时间段过度使用储能供电，导致在用电高峰时段储能设备处于过放状态不

能进行供电，在用电高峰时段 11-16 时从主电网购入了较多电量没有充分的利用可再生能源。

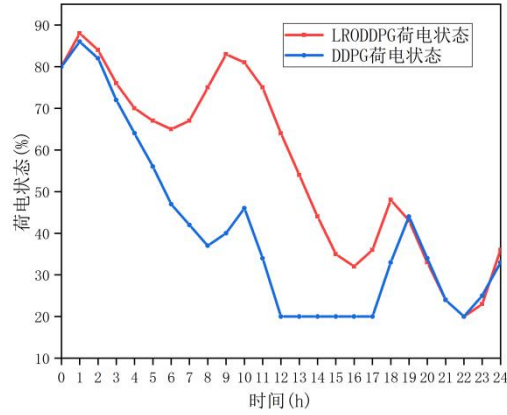


图 4-11 两种算法的储能电荷情况

图 4-11 为两种智能体在能量调度过程中储能电荷的变化状态，DDPG 智能体在调度过程中在用户负荷高峰时存在着过度放电的情况，导致储能设备陷入停滞，而 LRODDPG 智能体可以有效规避过度放电问题，让储能设备一直参与能量调度过程。

表 4-6 优化效果对比

	LRODDPG	DDPG
成本(\$)	69.3	269.6
自平衡率(%)	99.25%	96.87%
碳排放量(吨)	4.35	4.01

表 4-6 展示了两两种算法在典型日的优化效果对比。从运行成本来看 LRODDPG 智能体的日均成本为 69.3 美元而 DDPG 智能体的日均成本为 269.6 美元,降低了 74.29%;而从自平衡角度分析两个智能体的自平衡率分别为 99.25% 和 96.87%, 降低了 2.38%; 同时优化后的算法降低了 4.35 吨的碳排放量。

综合来看经过优化后 LRODDPG 算法成本更低、自平衡率更高。在经济和技术性能方面都有着更好性能。

4.5 本章小结

在本章中提出了一种的基于 LRODDPG 算法的微电网能量管理方案。首先使用长短期记忆神经网络（LSTM）对源荷时序数据特征提取得到有效的预测模

型解决分布式能源及负荷不确定的问题；其次对奖励函数进行优化，将奖励分解为即时奖励和完成奖励，提高了智能体的训练效率。最后通过仿真实验分析对比优化算法的和常规算法的收敛速度、实际调度情况以及成本、自平衡率控制效果，验证了 LRODDPG 算法可以更快更好的完成微电网的能量优化管理。

第 5 章 基于 BMADDPG 的多微电网双层能量管理

5.1 引言

在上一章中对单一微电网进行了基于深度强化学习的微电网能量管理，但随着微电网技术的不断的发展和进步以及可再生能源的利用率不断提高，单一微电网存在可靠性不足、负载平衡困难、能源利用效率低等问题逐渐显现。为了解决这些问题，发展多微电网架构成为一种有前景的研究方向，通过多个微电网之间的协调和能量交互，提高系统的可靠性、灵活性、能源利用效率和安全性^[59]。

相比于单微电网，多微电网的优势可以概括为以下几点：第一可以提高可靠性，多微电网具有分布式能源资源和负荷需求，其中的每个微电网相互连接，形成一个网络。第二可以提高经济性，多微电网可以实现能源的共享与优化调度，提高能源利用效率，从而降低整体能源成本；第三增强整个系统的灵活性和适应性，多微电网可以更好地适应能源需求的变化和负荷波动。第四对可持续性有所提高，多微电网的结构使得可再生能源的利用更加可持续和高效^[60]。综上所述，从可靠性、经济性、灵活性和未来的发展趋势来看，对多微电网的能量管理研究有着很大的实际意义。

5.2 多微电网双层能量管理模型

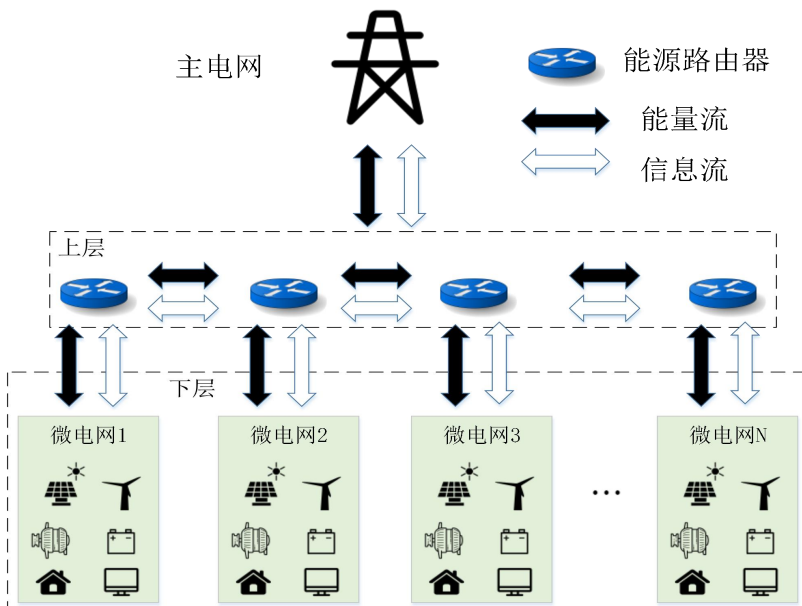


图 5-1 多微电网双层结构图

在本章中采用多微电网双层能量管理模型如图 5-1 所示, 下层由 N 个微电网构成, 多个微电网通过能源路由器相连, 能源路由器在多微电网中起到了实现能源交互、调度和优化的关键作用。上层是根据下层抽象出的拓扑图, 方便进行潮流分析。在进行能源管理的过程中, 下层首先使用深度强化学习算法对可控设备的出力进行决策, 随后上层根据下层调控的结果计算最优潮流优化, 通过双重优化后计算奖励值并更新网络参数以获得更好的效果。

5.2.1 下层多智能体深度强化学习算法

多智能体深度确定性策略梯度算法(MADDPG)是 Deep Mind 的研究人员基于 DDPG 算法和 actor-critic 框架进行改进所提出的一种多智能体深度强化学习算法。MADDPG 算法结构如图 5-2 所示, 共有 n 个智能体, 每个智能体都有独立的 Critic 网络和 Actor 网络。在训练过程中, Critic 网络将所有智能体的全局观测 $(O_1, O_2, \dots, O_n)$ 和联合动作 $(a_1, a_2, \dots, a_n)$ 作为输入, 输出该环境下动作的 Q 值, 并根据 Q 值更新网络参数。而在测试过程中, 各智能体只需要自己的观测就可以输出动作策略, 即统一训练, 独立执行。

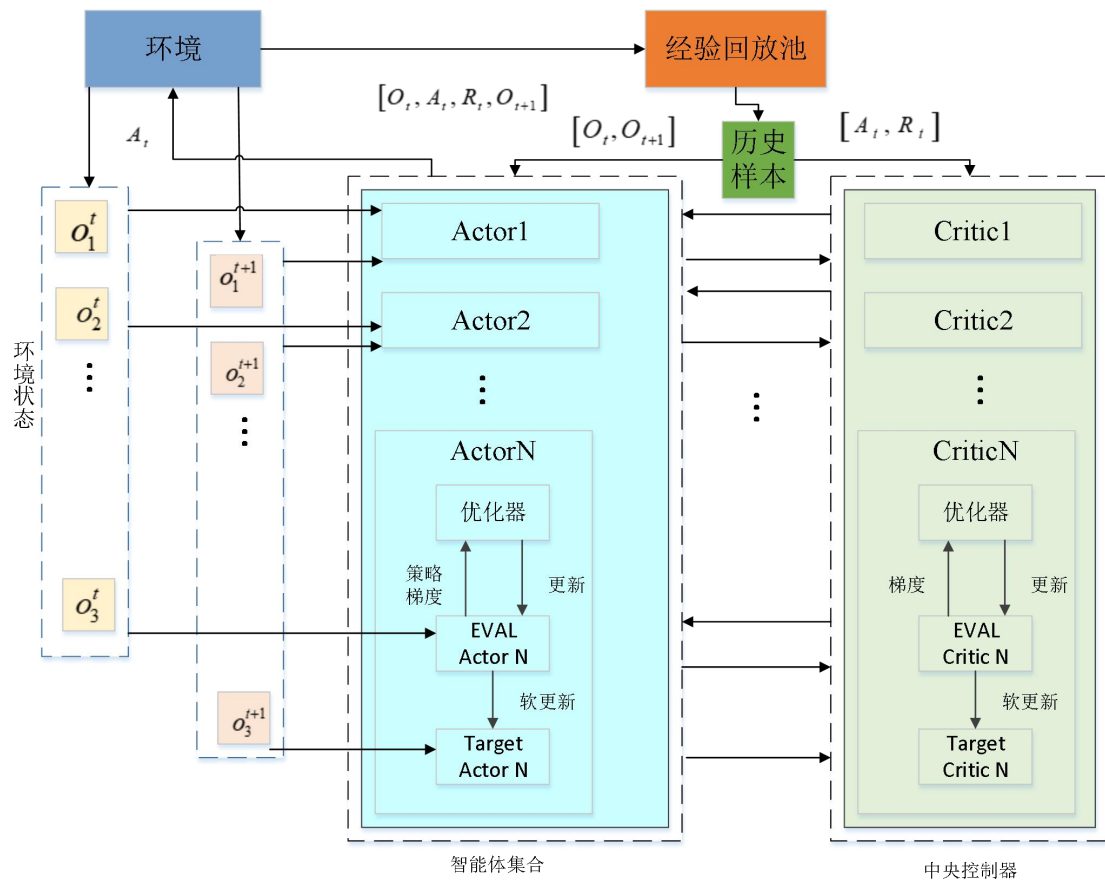


图 5-2 MADDPG 算法结构图

网络参数的具体更新方式如下：

利用 $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ 代表 n 个智能体策略函数的参数，用 $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ 表示 n 个智能体的策略函数。针对第 i 个智能体，把累计奖励期望值定义为：

$$J(\theta) = \mathbb{E}[R_i] = \mathbb{E}_{s \sim p^\pi, a_i \sim \pi_{\theta_i}} \left[\sum_{t=0}^{\infty} \gamma^t r_i, t \right] \quad (5-1)$$

式中 r_i 表示第 i 个智能体的奖励， γ 表示折扣率。

针对随机策略梯度，求解策略梯度的公式为：

$$\nabla_{\theta_i} J(\theta_i) = \mathbb{E}_{s \sim p^\mu, a_i \sim \pi_i} \left[\nabla_{\theta_i} \log \pi_i(a_i | o_i) Q_i^\pi(x, a_1, \dots, a_n) \right] \quad (5-2)$$

式中 $Q_i^\pi(x, a_1, \dots, a_n)$ 是第 i 个智能体的评价函数，它的输入是每个智能体采取的动作 a_i 以及环境信息 x ，输出为第 i 个智能体的 Q 值。

如果将随机策略替换为确定性的策略，用 μ_i 表示第 i 个智能体的策略，则第 i 个智能体的梯度可以写成：

$$\nabla_{\theta_i} J(\mu_i) = \mathbb{E}_{s, a \sim D} \left[\nabla_{\theta_i} \mu_i(a_i | o_i) Q_i^\mu(x, a_1, \dots, a_n) \Big|_{a_i = \mu_i(o_i)} \right] \quad (5-3)$$

中心化的动作值函数可以通过反向传播算法进行更新：

$$L(\theta_i) = \mathbb{E}_{x, a, r, x'} \left[\left(Q_i^\mu(x, a_1, \dots, a_n) - y \right)^2 \right] \quad (5-4)$$

$$y = r_i + \gamma Q_i'(x^i, a_1^i, \dots, a_n^i) \Big|_{a_j^i = \mu_j^i(o_j)} \quad (5-5)$$

5.2.2 上层最优潮流模型

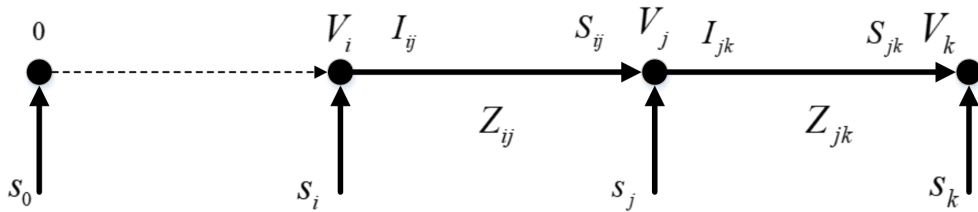


图 5-3 支路潮流结构图

目前配电网运行模式主要包括：辐射状运行、环状运行、网状运行和分布式运行等，本文研究的微电网采用辐射状运行模式，支路潮流法适用于辐射状的网络，所以本章使用支路潮流模型对多微电网进行建模。图 5-3 为支路潮流结构图，图中， V_k 为节点电压， I_{jk} 为支路电流，支路 jk 表示从节点 j 到节点 k 为潮流正

方向 $\delta(j)$ 为以 j 为首端节点的支路末端节点集合, $\pi(j)$ 为以 j 为末端节点的支路首端节点集合。 $s_k = p_k + iq_k$ 为节点 k 的注入功率, $S_{jk} = P_{jk} + iQ_{jk}$ 为支路的流经功率, $Z_{jk} = r_{jk} + ix_{jk}$ 为支路阻抗。集合 A 表示所有节点集合, E 表示所有支路集合。根据 Fu L 等人的研究^[61], 最优潮流的一般模型如下式所示:

$$\min f(p, q, P, Q, V, I) \quad (5-6)$$

$$\text{s.t.} \begin{cases} p_j = \sum_{k \in \delta(j)} P_{jk} - \sum_{i \in \pi(j)} (P_{ij} - I_{ij}^2 r_{ij}) + g_j V_j^2, \quad \forall j \in A \\ q_j = \sum_{k \in \delta(j)} Q_{jk} - \sum_{i \in \pi(j)} (Q_{ij} - I_{ij}^2 x_{ij}) + b_j V_j^2, \quad \forall j \in A \end{cases} \quad (5-7)$$

$$V_j^2 = V_i^2 - 2(P_{ij} r_{ij} + Q_{ij} x_{ij}) + I_{ij}^2 (r_{ij}^2 + x_{ij}^2), \quad \forall ij \in E \quad (5-8)$$

$$I_{ij}^2 = \frac{P_{ij}^2 + Q_{ij}^2}{V_i^2}, \quad \forall ij \in E \quad (5-9)$$

$$\underline{I}_{ij} \leq I_{ij} \leq \bar{I}_{ij}, \quad \forall ij \in E \quad (5-10)$$

$$\underline{V}_j \leq V_j \leq \bar{V}_j, \quad \forall j \in A \quad (5-11)$$

式 (5-6) 为目标函数, 可以为发电成本最小、环境成本最小、运行风险最小等; 式 (5-7) ~ (5-11) 为潮流平衡约束和安全约束。

但是该最优潮流模型的非线性规划模型, 由于非线性规划模型复杂性和数值稳定性问题, 在计算时间和资源需求量都会更高, 迭代过程中的发散收敛速度也会变慢, 因此需要对式 (5-9) 进行 SORC 转化。SORC 方法在求解最优潮流问题时, 通过对系统的约束条件进行逐步放宽, 从而降低问题的非线性程度, 使其更易于求解。令 $\tilde{I}_{ij} = I_{ij}^2$ 、 $\tilde{V}_j = V_j^2$, 再对其进行求解即可得到全局最优解。转化结果为:

$$\min f \left(p, q, P, Q, \tilde{V}, \tilde{I} \right) \quad (5-12)$$

$$\text{s.t.} \begin{cases} p_j = \sum_{k \in \delta(j)} P_{jk} - \sum_{i \in \pi(j)} \left(P_{ij} - \tilde{I}_{ij} r_{ij} \right) + g_j \tilde{V}_j, \quad \forall j \in A \\ q_j = \sum_{k \in \delta(j)} Q_{jk} - \sum_{i \in \pi(j)} \left(Q_{ij} - \tilde{I}_{ij} x_{ij} \right) + b_j \tilde{V}_j, \quad \forall j \in A \end{cases} \quad (5-13)$$

$$\tilde{V}_j = \tilde{V}_i - 2(P_{ij} r_{ij} + Q_{ij} x_{ij}) + \tilde{I}_{ij} (r_{ij}^2 + x_{ij}^2), \quad \forall ij \in E \quad (5-14)$$

$$\left\| \begin{array}{c} 2P_{ij} \\ 2Q_{ij} \\ \tilde{I}_{ij} - \tilde{V}_j \end{array} \right\|_2 \leq \tilde{I}_{ij} + \tilde{V}_j, \quad \forall ij \in E \quad (5-15)$$

$$\underline{I}_{ij}^2 \leq \tilde{I}_{ij} \leq \bar{I}_{ij}^2, \quad \forall ij \in E \quad (5-16)$$

$$\underline{V}_j^2 \leq \tilde{V}_j \leq \bar{V}_j^2, \quad \forall j \in A \quad (5-17)$$

式 (5-12) 至 (5-17) 为松弛后的最优潮流模型, 通过松弛这些约束, 可以将非凸问题转化为凸问题, 降低算法的求解难度方便设计更有效的求解策略, 保证找到全局最优解, 同时现有的商业优化求解器大多数擅长解决凸优化问题, 为后续的方案奠定了基础。

5.3 基于 BMADDPG 的多微电网双层能量管理方案

5.3.1 约束条件

(1) 支路潮流模型约束:

$$\text{s.t.} \begin{cases} p_j = \sum_{k \in \delta(j)} P_{jk} - \sum_{i \in \pi(j)} \left(P_{ij} - \tilde{I}_{ij} r_{ij} \right) + g_j \tilde{V}_j, \quad \forall j \in A \\ q_j = \sum_{k \in \delta(j)} Q_{jk} - \sum_{i \in \pi(j)} \left(Q_{ij} - \tilde{I}_{ij} x_{ij} \right) + b_j \tilde{V}_j, \quad \forall j \in A \end{cases} \quad (5-18)$$

$$\tilde{V}_j = \tilde{V}_i - 2(P_{ij} r_{ij} + Q_{ij} x_{ij}) + \tilde{I}_{ij} (r_{ij}^2 + x_{ij}^2), \quad \forall ij \in E \quad (5-19)$$

$$\left\| \begin{array}{c} 2P_{ij} \\ 2Q_{ij} \\ \tilde{I}_{ij} - \tilde{V}_j \end{array} \right\|_2 \leq \tilde{I}_{ij} + \tilde{V}_j, \quad \forall ij \in E \quad (5-20)$$

$$\underline{I}_{ij}^2 \leq \tilde{I}_{ij} \leq \bar{I}_{ij}^2, \quad \forall ij \in E \quad (5-21)$$

$$\underline{V}_j^2 \leq \tilde{V}_j \leq \bar{V}_j^2, \quad \forall j \in A \quad (5-22)$$

式中的参数在 5.2.2 章节介绍在此不再重复介绍。

(2) 功率平衡约束:

$$P_{m,t}^{pv} + P_{m,t}^{WT} + P_{m,t}^{ESS} + P_{m,t}^{MG} + P_{m,t}^{MT} - P_{m,t}^L = p_{m,t}, \forall m \in N \quad (5-23)$$

$$Q_{m,t}^{MG} - Q_{m,t}^L = q_{m,t}, \forall m \in N \quad (5-24)$$

式(5-23)为多微电网的有功功率平衡约束, $P_{m,t}^L$ 为 t 时刻的负荷功率(kW), $p_{m,t}$ 为 t 时刻的注入有功功率(kVar); 式(4-23)为无功功率平衡约束, $Q_{m,t}^{MG}$ 、 $Q_{m,t}^L$ 分别为 t 时刻微电网和负载的无功功率(kVar), $q_{m,t}$ 为 t 时刻的注入无功功率(kVar)。

(3)燃气轮机约束:

$$P_{m,\min}^{MT} < P_{m,t}^{MT} < P_{m,\max}^{MT}, \forall m \in N \quad (5-25)$$

式中 $P_{m,t}^{MT}$ 为 t 时刻燃气轮机的发电功率(kW)。

(4)主电网功率约束:

$$P_{m,\min}^{MG} < P_{m,t}^{MG} < P_{m,\max}^{MG}, \forall m \in N \quad (5-26)$$

$$Q_{m,\min}^{MG} < Q_{m,t}^{MG} < Q_{m,\max}^{MG}, \forall m \in N \quad (5-27)$$

式中 $Q_{m,t}^{MG}$ 为主电网在 t 时刻的无功功率(kVar)。

(5)储能电池约束:

$$P_{m,\min}^{ESS} < P_{m,t}^{ESS} < P_{m,\max}^{ESS}, \quad \forall m \in N \quad (5-28)$$

$$SOC_{m,\min} \leq SOC_{m,t} \leq SOC_{m,\max} \quad \forall m \in N \quad (5-29)$$

式中 $P_{m,t}^{ESS}$ 、 $SOC_{m,t}$ 分别为 t 时刻储能电池的充放电功率(kW)和荷电状态。

5.3.2 优化目标函数

在本节中多微电网优化目标为最小化主电网交易成本 $f_{m,t}^{MG}$ 、设备维护成本 $f_{m,t}^{EM}$ 和燃气轮机运行成本 $f_{m,t}^{MT}$, m 表示在多微电网中的第 m 个微电网; 目标函数为:

$$\min f_t = \min \sum_{t=1}^T (f_{m,t}^{MG} + f_{m,t}^{EM} + f_{m,t}^{MT}) \quad (5-30)$$

(1)主电网的运行成本为:

$$f_{m,t}^{MG} = P_{m,t}^{MG} c_t^{MG} \Delta t \quad (5-31)$$

$P_{m,t}^{MG}$ 为在 t 时刻从主电网购电的有功功率值(kW); c_t^{MG} 为主电网在 t 时刻的实时电价(\$/kWh)。

(2)光伏、风力设备和储能设备成本:

$$f_{m,t}^{EM} = [K_{PV} P_{m,t}^{PV} + K_{WT} P_{m,t}^{WT} + K_{ESS} (P_{m,t}^{ESSch} + P_{m,t}^{ESSdis})] \Delta t \quad (5-32)$$

式中 K_{PV} 、 K_{WT} 和 K_{ESS} 分别为光伏、风力发电设备和储能设备的运行维护成本系数； $P_{m,t}^{PV}$ 和 $P_{m,t}^{WT}$ 分别为 t 时刻光伏、风力发电功率； $P_{m,t}^{ESSch}$ 和 $P_{m,t}^{ESSdis}$ 分别为储能设备在 t 时刻的充放电功率。

(3) 燃气轮机运行成本：

$$f_{m,t}^{MT} = \left(\frac{c_t^{gas}}{H^{ng}} \frac{P_{m,t}^{MT}}{\xi^{MT}} + K_{MT} P_{m,t}^{MT} \right) \Delta t \quad (5-33)$$

式中 c_t^{gas} 为 t 时刻的天然气实时价格 ($\$/m^3$)， H^{ng} 为天然气热值 (kWh/Nm^3)， ξ^{MT} 为燃气轮机的发电效率 (%)， K_{MT} 为燃气轮机的运行维护成本系数， $P_{m,t}^{MT}$ 为 t 时刻的燃气轮机运行功率。

(4) 网络线损：

$$f_t^{LOSS} = \sum_i^E I_{ij,t} r_{ij} \rho_t \quad (5-34)$$

式中 $I_{ij,t}$ 为 t 时刻的第 ij 条支路电流 (A)， r_{ij} 为第 ij 条支路阻抗。

5.3.3 构建马尔可夫决策过程

在第四章对单一微电网进行能量管理时构建了马尔可夫决策模型，在本章使用 MADDPG 深度强化学习算法对多微电网进行能量管理时，首先也需要让神经网络模型明确控制目标和决策流程，即将多微电网能量管理问题转化为马尔可夫决策过程：

(1) 状态空间：

$$S_t = \begin{bmatrix} P_{1,t}^{PV}, P_{1,t}^{WT}, P_{1,t}^L, c_t^{MG}, SOC_{1,t} \\ \vdots \\ P_{m,t}^{PV}, P_{m,t}^{WT}, P_{m,t}^L, c_t^{MG}, SOC_{m,t} \end{bmatrix} \quad (5-35)$$

与单一微电网状态空间不同的是，多微电网的 Critic 网络需要接受多微电网中所有单一微电网的状态信息，所以式中包含了从 1 到 m 个微电网的所有状态信息；式中 $P_{m,t}^{PV}$ 、 $P_{m,t}^{WT}$ 、 $P_{m,t}^L$ 、 c_t^{MG} 和 $SOC_{m,t}$ 分别为 t 时刻的光伏、风力发电功率、负载功率、实时电价和储能电池荷电状态。

(2) 动作空间：

$$A_t = \begin{bmatrix} P_{1,t}^{MG}, P_{1,t}^{ESS}, P_{1,t}^{MT}, Q_{1,t}^{MG} \\ \vdots \\ P_{m,t}^{MG}, P_{m,t}^{ESS}, P_{m,t}^{MT}, Q_{m,t}^{MG} \end{bmatrix} \quad (5-36)$$

式中 $P_{m,t}^{MG}$ 、 $P_{m,t}^{ESS}$ 、 $P_{m,t}^{MT}$ 和 $Q_{m,t}^{MG}$ 分别为 t 时刻的主电网交换的有功功率、储能电池的充放电功率、微型燃气轮机的发电功率和主电网交换的无功功率。

(3) 奖励函数:

双层能量管理模型包含了智能体的决策阶段以及最优潮流计算阶段,所以在奖励函数的设置上也需要对这两个阶段进行分别设计,下面为具体设计方案:

$$r_{m,t} = - \left((\alpha \cdot f_t) + \beta \cdot \sum_{i=1}^N \text{cons}(SOC_{i,t}) \right) \quad (5-37)$$

$$f_t = f_{m,t}^{MG} + f_{m,t}^{EM} + f_{m,t}^{MT} \quad (5-38)$$

$$\text{cons}(x) = \ln \frac{|x - x_{\max}| + |x - x_{\min}|}{(x_{\max} - x_{\min})} \quad (5-39)$$

上式中 $r_{m,t}$ 为智能体的奖励函数; α 和 β 为奖励系数,可以通过调整奖励系数大小确定控制目标的重要性, $\text{cons}(x)$ 为惩罚系数,控制目标为储能设备的电荷状态,在储能电荷过高或过低时执行惩罚,训练智能体在将储能电荷保持在较为良好的状态避免过充和过放状况的发生。

$$r_{tfc} = - \left((\gamma \cdot f_t^{LOSS}) + \lambda |Q_{m,Real}^{ESS} - Q_{m,Esti}^{ESS}| \right) \quad (5-40)$$

上式中 r_{tfc} 为最优潮流计算的奖励函数;在最优潮流计算阶段,奖励函数的设置目标是在正常运行的情况下尽量降低有功功率损耗 f_t^{LOSS} ,同时为了加快对智能体控制策略的掌握速度,在奖励函数的设计上还加入了指导机制,即在奖励函数加入 $\lambda |Q_{m,Real}^{ESS} - Q_{m,Esti}^{ESS}|$,其中 $Q_{m,Real}^{ESS}$ 、 $Q_{m,Esti}^{ESS}$ 分别为储能系统的充放电功率的实际值和预估值, γ 、 λ 为奖励系数。

$$R_T = \sum_j^n r_{tfc} + \sum_{m=1}^N r_{m,t} \quad (5-41)$$

R_T 为最终双层能量管理模型的奖励函数,包含了智能体和潮流计算两部分的奖励值,为了简化分析过程,本章没有使用第四章提出的奖励优化函数。

5.3.4 MADDPG 算法流程

BMADDPG 算法流程如算法 5-1 所示：

算法 5-1 BMADDPG 算法

输入：初始状态

输出：能量管理策略

初始化：Actor 网络参数、Critic 网络参数、经验池 D

1: **for** episod=1 to M **do**

2: **for** t=1 to T **do**

3: 步骤一：下层多微电网智能体输出动作 $a_{m,t}$ ，得到优化变量

$$\{P_{m,t}^{MT}, P_{m,t}^{ESS}, P_{m,t}^{MG}, Q_{m,t}^{MG}\}, m \in N$$

4: 步骤二：上层使用商业优化解决最优潮流问题，得到优化变量

$$\{v_{j,t}\}, \{p_{j,t}, q_{j,t}\}, \{l_{ij,t}\}, \{P_{ij,t}, Q_{ij,t}\}, \{\Delta P_{m,t}^{MG}, \Delta Q_{m,t}^{MG}\}$$

5: 步骤三：将得到的优化变量用于多微电网能量管理，获得奖励 $r_{m,t}$ ，进

入下个状态 $s_{m,t+1}$ ，将 (s_t, a_t, r_t, s_{t+1}) 存入经验池 D

6: **end for**

7: **for** t=1 to T **do**

8: 从经验池 D 中随机采样一批数据：

$$y = r_i + \gamma Q_i'(x^i, a_1^i, \dots, a_n^i) \big|_{a_j^i = \mu_j^i(o_j)}$$

9: 通过损失函数最小更新 Critic 网络参数：

$$L(\theta_i) = \mathbb{E}_{x,a,r,x'} \left[\left(Q_i^\mu(x, a_1, \dots, a_n) - y \right)^2 \right]$$

10: 通过梯度反向传播更新 Actor 网络参数：

$$\nabla_{\theta_i} J(\theta_i) = \mathbb{E}_{s \sim \rho^\mu, a_i \sim \pi_i} \left[\nabla_{\theta_i} \log \pi_i(a_i | o_i) Q_i^\pi(x, a_1, \dots, a_n) \right]$$

11: **end for**

12: 通过软更新更新每个网络的参数

13: **end for**

在智能体决策阶段结束后，可能会存在实际功率平衡条件与调度结果不匹配的情况，为了解决这一问题，在进行潮流优化时还需要考虑 $\Delta P_{m,t}^{MG}$ 、 $\Delta Q_{m,t}^{MG}$ ，分别为 t 时刻主电网的有功功率和无功功率的改变量，通过调整主电网的功率可以

保证满足功率平衡的条件。

5.4 实验结果分析

5.4.1 仿真实验数据

本文多微电网由三个独立微电网组成，将 IEEE33 节点系统分解为三个微电网区域，在本章中选取 1、7、21 为连接节点，形成包含三个微电网的多微电网结构。各微电网的设备容量及参数如表 5-1 所示，储能设备的相关参数如表 5-2 所示。本章所使用的可再生能源发电数据、负荷数据和实时电价数据分别来自开源数据平台 Open Power System Data。

表 5-1 光伏、风力、燃气轮机设备参数设置

设备类型	微电网序号	最小容量/kW	最大容量/kW
光伏设备	1	0	600
	2	0	300
	3	0	600
风力设备	1	0	500
	2	0	300
	3	0	500
微型燃气轮机	1	0	300
	2	0	200
	3	0	300

表 5-2 储能设备参数设置

	微电网序号	容量配置 /kWh	充放电功率 /%	SOC_{min} /%	SOC_{max} /%
储能设备	1	600	95	20	85
	2	300	95	20	85
	3	600	95	20	85

仿真实验所用的设备处理器为 AMD Ryzen7 5800H, 显卡为 NVIDIA GeForce RTX3050, 在 Python 3.6 和 TensorFlow 框架下编写程序, 进行仿真实验。使用基于 Python 的电力系统分析工具库 PYPOWER 求解最优潮流问题, 多智能体深度强化学习框架为 Pytoch,MADDPG 参数设置为: 学习率设置为 0.005, 折扣因子设置为 0.95, 包含两个隐藏层, 每个隐藏层神经元数为 128, 训练轮数设置为

10000。

5.4.2 训练效果分析

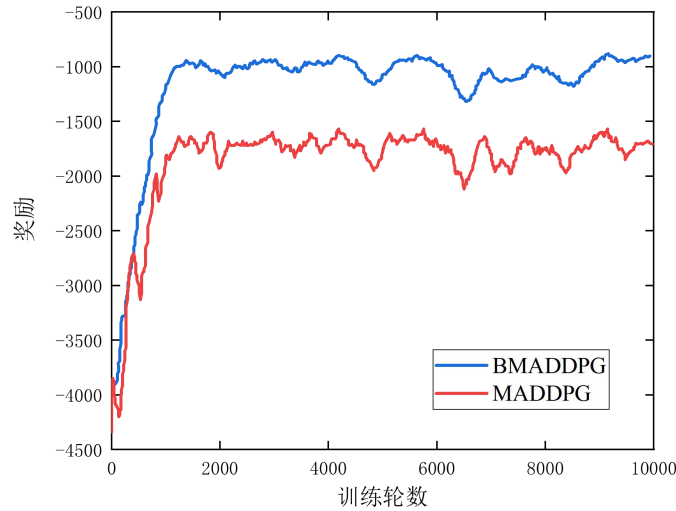


图 5-4 BMADDPG 和 MADDPG 算法训练过程

图 5-4 为使用 BMADDPG 双层能量管理方案的奖励值和只采用 MADDPG 深度强化学习算法的训练过程对比。从宏观角度来看两种方法的奖励曲线均呈现上升趋势，说明两种方法均可以完成对多微电网目标的能量管理目标。采用 BMADDPG 的奖励曲线约在 2000 集后逐渐收敛于-1000 左右，且在训练结束前仍呈现上升趋势；而无潮流计算的奖励曲线约在 3000 集左右开始逐渐收敛于-1700 左右，在训练后期的奖励波动较大，且训练结果最后呈现下降趋势，表明该方法并没有取得最优解；在训练前期两种算法也有着一定差异，无潮流计算的奖励曲线上升过程中的波动较大且速度较为缓慢，加入潮流计算的奖励曲线更为平滑。

通过对比可以得出结论：加入潮流计算的双层能量管理方案在解决多微电网能量调度问题相比传统只使用深度强化学习算法的方案具有更大优势。

5.4.3 调度结果分析

仅仅通过算法收敛获得奖励的方法不能直观的展现出算法的实际性能，所以将训练好的智能体选取随机一个微电网典型日进行决策。由于微电网 1 与微电网 3 的设置参数相同，本节中选取 MG1 与 MG2 进行典型日的调度结果进行分析。

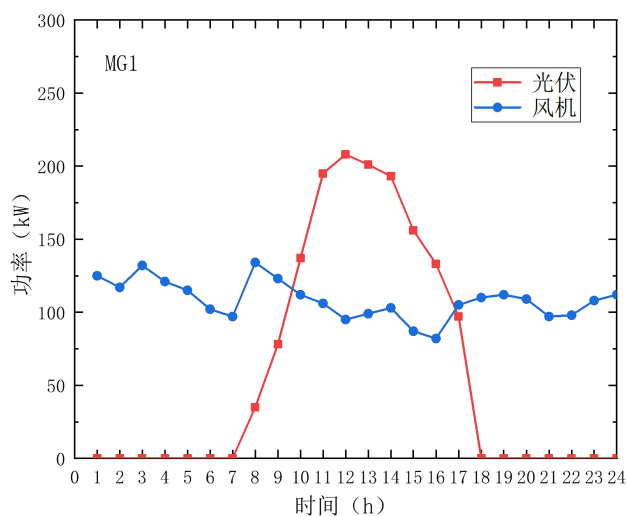


图 5-5 MG1 的可再生能源出力情况

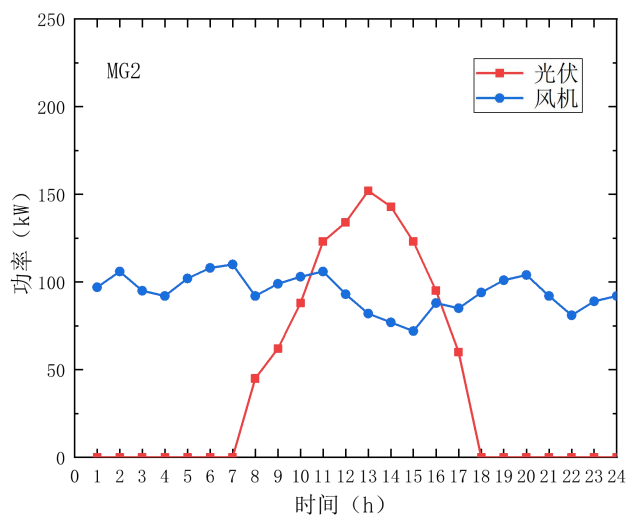


图 5-6 MG2 的可再生能源出力情况

图 5-5、5-6 分别为 MG1 和 MG2 在典型日内可再生能源出力情况，两个微电网的大致趋势相同，光伏发电时间集中在 7-18 时，风力发电在全天都有出力。

图 5-7、5-8 分别为 MG1 和 MG2 的用户负荷曲线、储能设备的充放电功率、微型燃气轮机运行功率以及主电网交易功率的柱状结合图。通过对比两个图可以发现负荷功率在 8-20 时处于较高状态，凌晨和深夜的负荷功率较低，结合实时电价进行分析，负荷功率的大小与电价高低成正比例关系，所以根据实时电价对每个微电网的可控设备进行调度，降低从主电网的购电量，完成了多微电网能量管理的核心目标。

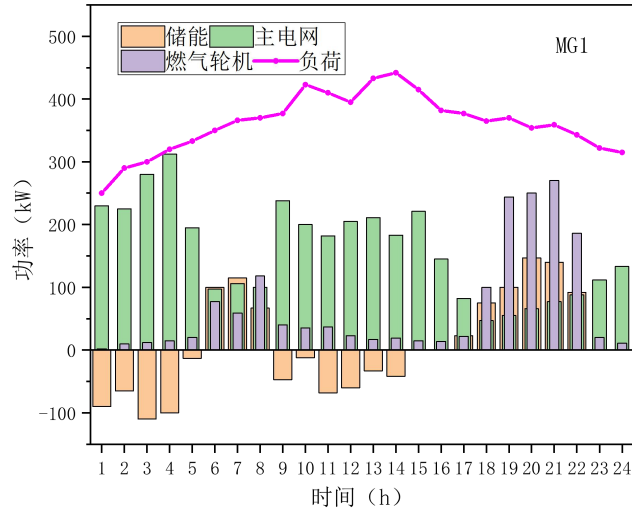


图 5-7 MG1 的可控设备出力情况

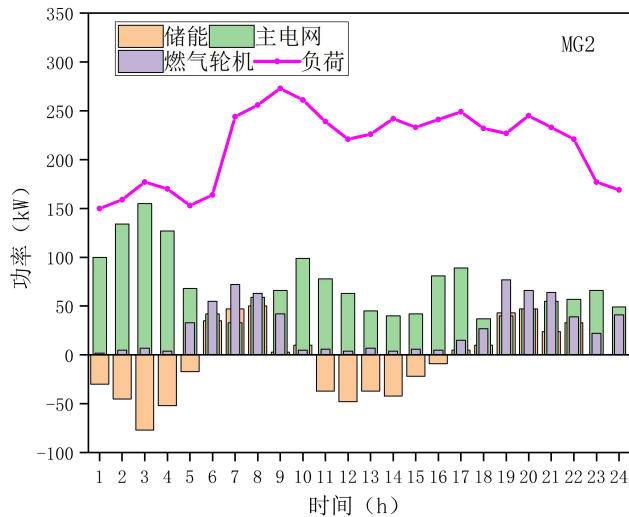


图 5-8 MG2 的可控设备出力情况

在电价较低的凌晨时段(1-5 时)，电价较低从主电网购买了大量电力，微型燃气轮机此时也处于低功率的运行，储能装置在此阶段进行充电为负荷高峰时段做好准备；在负荷较高且电价较高时段(6-8 时、18-21 时)，储能装置均处在放电状态，同时微型燃气轮机也处在高功率运行，两种设备共同出力弥补了一定的负荷功率缺口，降低了从主电网的购电量，可以发现在这段时间主电网的交易功率也处在全天中的较低阶段。而在电价适中的时段(9-17 时)，储能设备和微型燃气轮机的运行状态与凌晨时段类似。

通过上述分析可以得到：基于 BMADDPG 的多微电网双层能量管理方案可

以有效降低微电网的运行成本。

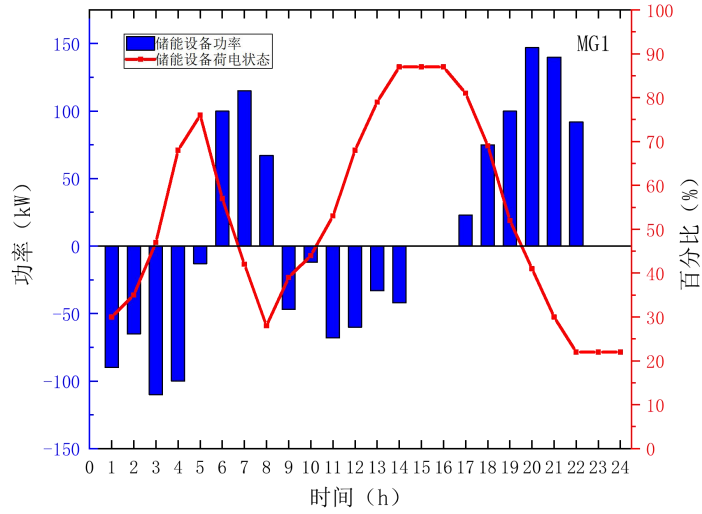


图 5-9 MG1 的储能设备功率及荷电状态

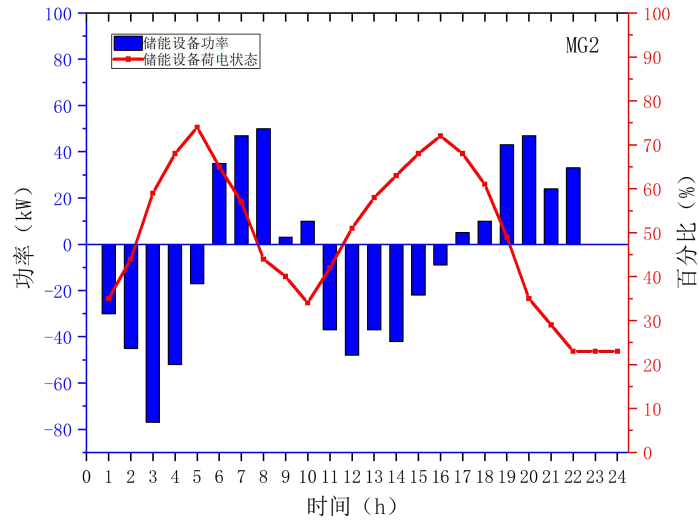


图 5-10 MG2 的储能设备功率及荷电状态

图 5-9、5-10 分别为 MG1 和 MG2 的储能设备功率值和荷电状态，在储能设备功率为负值时表明储能电池处于充电状态，荷电状态上升，反之荷电状态下降。设定两个微电网荷电状态的上下限为 85%和 20%，超过上限状态储能设备处于过充状态而低于下限则储能设备处于过放状态。通过分析两张图片可以得到：两个微电网在大部分时间荷电状态均处于安全状态，只有 MG1 在 14-16 时出现了短暂的过充现象，过充比例约为 3%。

通过上述分析可以得到：基于 MADDPG 深度强化学习算法的多微电网复合

能量管理模型可以有效的控制储能设备的运行，将荷电状态稳定在安全状态下。

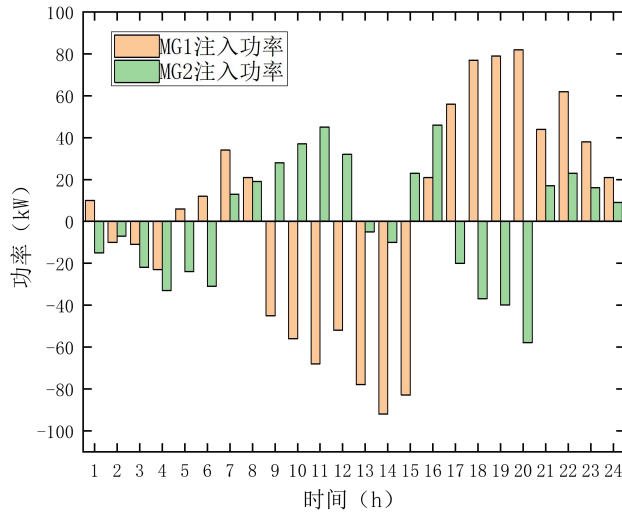


图 5-11 MG1、MG2 的上层优化结果

图 5-11 为 MG1 和 MG2 的上层优化结果柱状图，展示了两个微电网的注入功率。在多微电网系统中每个微电网节点可以向系统中注入电功率，节点注入功率的大小和方向会影响整个多微电网的运行状态和稳定性。当节点注入功率为正时，节点会将功率传输至电网中，反之则会从电网中吸收功率保证微电网的功率平衡。本章提出的基于 BMADDPG 的双层能量管理方案加入了潮流优化环节，可以让节点的注入功率在微电网中相互传输，如在 8-12 时 MG2 的注入功率为正而 MG1 的注入功率为负，MG2 即可从 MG1 中获得功率完成互济目标，保证功率平衡。

5.4.4 不同方案效果对比

为了进一步验证 MADDPG 算法的有效性，将进行对比试验，对比多种算法的决策时间以及运行成本。对比试验的方案包括：

方案一：MADDPG(Multi-Agent Deep Deterministic Policy Gradient)，使用单层能量管理方案，只使用 MADDPG 算法对多微电网进行能量管理。

方案二：BCOMA(Bilayer Cooperative Multi-Agent Learning with Communication)，使用双层能量管理方案，下层使用 COMA 深度强化学习算法。

方案三：BMAPPO(Bilayer Multi-Agent Proximal Policy Optimization)，使用双层能量管理方案，下层使用 MAPPO 深度强化学习算法。

方案四：BMPC(Bilayer Model Predictive Control)使用双层能量管理方案，

下层使用 MPC 算法求解。

表 5-3 不同方案的执行时间和成本对比

	BMADDPG	MADDPG	BCOMA	BMAPPO	BMPC
执行时间 (s)	0.512	0.277	0.525	0.586	62.54
运行成本 (\$)	591.88	658.26	622.54	612.43	635.86

通过表 5-3 的对比可以发现：从在线执行时间来看 MADDPG 由于只进行了下层能量优化，所以在线执行时间最短为 0.277 秒，其余三组深度强化学习算法时间分别为 0.512 秒、0.525 秒和 0.586 秒均较为接近，可以分析得到基于深度强化学习算法可以满足实时性的需求。而下层使用 MPC 算法的在线决策时间为 62.54 秒，相较于使用深度强化学习算法时间大大增加，很难满足实时性的要求。

从运行成本角度分析：首先使用基于 BMADDPG 的双层能量管理方案的运行成本相较于仅使用 MADDPG 的单层能量管理方案成本降低了 10.08%，证明了本章所提出的双层能量管理方案能够有效降低运行成本。其次横向对比使用双层能量管理四种方案，BMADDPG 算法相较于 BCOMA、BMAPPO 和 BMPC 算法成本分别降低了 4.92%、3.36%和 6.92%。说明了本章所使用的算法相较于其他算法的优化效果更好，可以更有效地处理多微电网能量管理问题。

5.5 本章小结

本章针对多微电网设计了一种基于 BMADDPG 的多微电网双层能量管理方案，该方案下层首先使用多智能体深度确定性策略梯度算法对微电网设备出力进行决策，上层再对决策的结果进行最优化潮流；实验结果表明，本章所使用设计的 BMADDPG 能量管理方案可以最小化多微电网的运行成本，并且实现微电网间的功率互济。

第 6 章 总结与展望

6.1 总结

本文针对基于深度强化学习的微电网能量管理开展研究。在针对相对复杂的微电网环境时,相较于传统优化方法深度强化学习算法能够实现实时决策和快速计算,通过与环境不断交互执行动作并生成策略,解决微电网的能量管理问题。研究了单一微电网以及多微电网两种环境下的能量管理问题,并通过实验验证的管理方案的有效性。本文的主要工作内容和总结如下:

(1) 在深度强化学习的框架下构建了微电网能量管理模型,并使用深度强化学习算法在模型上训练智能体,实现了微电网的能量管理。

(2) 针对单一微电网能量管理问题,提出了基于 LRODDPG 的微电网能量管理方案。首先针对微电网中风电、光伏及负载不确定性较大的问题,利用 LSTM 神经网络提取环境中的时序数据未来趋势,将预测数据和外部状态空间输入智能体,改善连续调度空间下的算法收敛效果;随后对传统 DDPG 算法的奖励函数进行优化,将奖励函数分为即时奖励和完成奖励,两个奖励函数针对智能体在不同训练阶段给予不同的奖励;最后通过实验证明了本文提出的 LRODDPG 算法相较于传统 DDPG 算法,在算法收敛性、能源管理合理性还是在降低运行成本、提高自平衡率四个方面都具有较大提升,证明了基于 LRODDPG 的微电网能量管理方案的有效性。

(3) 针对多微电网环境,提出了基于 BMADDPG 的多微电网双层能量管理方案。多智能体强化学习方法在统一训练,独立执行的思想下解决多个智能体的协作决策问题,适用于多微电网环境。下层智能体对多微电网中的设备出力进行决策,随后上层根据调度结果再进行潮流优化,通过双重优化后计算奖励值并更新网络参数以获得更好的效果。在构建好双层能量管理模型以及马尔可夫决策过程后进行了仿真实验,实验从收敛速度和调度结果证明了方案的有效性;最后进行了对比试验,对比多种算法的运行时间以及运行成本,验证了加入潮流计算可以降低运行成本且保证微电网间的功率平衡。

6.2 展望

本文对单微电网以及多微电网分别设计了能量管理方案,并通过实验验证了方案的有效性,但随着电力系统的复杂化程度提高,智能化电网的趋势也越发明显,深度强化学习在微电网方面的应用也有着很大的提升空间,今后可以在以下两个方面进一步展开研究:

(1) 在大规模微电网中,博弈论可用于研究发电商和消费者之间的策略互动,以确定最佳的电力供应和需求平衡,在市场机制、需求响应、智能电网优化以及风险管理和决策中博弈论的加入都可以起到优化作用。而深度强化学习的不断发展也为解决博弈论问题提供了新的思路,如在多微电网环境中,每个微电网都可以被视为一个博弈参与者。后续可以通过将博弈论的概念和方法融入深度强化学习算法,使智能体可以学习到更加智能和适应性的策略,以在竞争或合作中取得更好的效果。

(2) 本文研究的单微电网和多微电网均与主电网相连,通过与主电网的交易可以保证微电网的正常运行,但随着不断的发展,偏远地区 and 海岛地区的孤岛微电网应用场景也越来越多,但孤岛微电网相比常规微电网,对可靠性和稳定性以及成本效益的设计要求更高,后续可以对孤岛微电网的能量管理进行研究。

参考文献

- [1] 潘家华, 陈梦玫, 刘保留. 净零碳转型的主要路径及其优化集成 [J]. 中国人口·资源与环境, 2023, 33(11): 1-12.
- [2] 李萌, 李雨珊, 潘家华等. 可再生能源、能源地缘政治风险与国家能源安全 [J]. 中国人口·资源与环境, 2022, 32(11): 1-8.
- [3] 陈逸文, 赵晋斌, 李军舟等. 电力低碳转型背景下氢储能的挑战与展望 [J]. 发电技术, 2023, 44(03):296-304.
- [4] Morais H, Pinto T, Vale Z. Adjacent Markets Influence Over Electricity Trading—Iberian Benchmark Study [J]. Energies. 2020,13(11):2808-2820.
- [5] Grünewald PH, Cockerill TT, Contestabile M, Pearson PJ. The socio-technical transition of distributed electricity storage into future networks—System value and stakeholder views [J]. Energy Policy. 2012 ,50: 49-57.
- [6] Yuan X C, Lyu Y J, Wang B, Liu Q H, Wu Q. China's energy transition strategy at the city level: The role of renewable energy [J]. Journal of Cleaner Production. 2018 , 205: 980-986.
- [7] 杨丽萍, 陆岷峰. “双碳”目标下数字绿色金融发展模式研究 [J]. 金融与经济, 2023, 43(12): 78-87.
- [8] Panda D, Turner O, Das S, et al. Prioritized experience replay based deep distributional reinforcement learning for battery operation in microgrids [J]. Journal of Cleaner Production, 2024, 434(3): 39-47.
- [9] Huang X, Zhang D, Zhang X. Energy management of intelligent building based on deep reinforced learning [J]. Alexandria Engineering Journal. 2021, 60(1):9-17.
- [10] Yu Z, Zhou B, Yang D, et al. Power System Operation Mode Calculation Based on Improved Deep Reinforcement Learning [J]. Mathematics, 2023, 12(1): 134-144.
- [11] Sayeed F, Hanumanthakari S, Oommen S, Anitha G, Hemavathi S, Pundir AK, Sudhakar M, Sathyamurthy R, Mohanavel V. A novel and comprehensive

- mechanism for the energy management of a Hybrid Micro-grid System [J] . Energy reports. 2022 , 8(3): 47-62.
- [12]Lipu M S H, Ansari S, Miah M S, et al. A review of controllers and optimizations based scheduling operation for battery energy storage system towards decarbonization in microgrid: Challenges and future directions [J] . Journal of Cleaner Production, 2022, 360(1): 1-30.
- [13]Dubey S M, Dubey H M, Salkuti S R. Modified quasi-opposition-based grey wolf optimization for mathematical and electrical benchmark problems [J] . Energies, 2022, 15(15): 5704-5723.
- [14]Garip I, Almulla A A, Hamza M S, et al. DG Compensation Using Optimal Distribution in Microgrids with High DG Penetration [J] . Electric Power Components and Systems, 2024, 11(1): 1-11.
- [15]Akulker H, Aydin E. Equipment Selection for Coupling a Microgrid with a Power-to-Gas System in the Context of Optimal Design and Operation [J] . Computers & Chemical Engineering, 2024, 181: 508-512.
- [16]Shuai H, Fang J, Ai X, et al. Stochastic optimization of economic dispatch for microgrid based on approximate dynamic programming [J] . IEEE Transactions on Smart Grid, 2018, 10(3): 2440-2452.
- [17]赵劲帅, 邱晓燕, 马菁曼等. 基于模糊聚类分析与模型识别的微电网多目标优化方法 [J] . 电网技术, 2016, 40(08): 2316-2323.
- [18]Suresh V, Janik P, Jasinski M, et al. Microgrid energy management using metaheuristic optimization algorithms [J] .Appl. Soft Comput. 2022, 134(9): 10-19.
- [19]Nemati M, Braun M, Tenbohlen S. Optimization of unit commitment and economic dispatch in microgrids based on genetic algorithm and mixed integer linear programming [J] . Applied energy, 2018, 210: 944-963.
- [20]李国庆, 翟晓娟, 李扬等. 基于改进蚁群算法的微电网多目标模糊优化运行 [J] .太阳能学报, 2018, 39(08): 2310-2317.
- [21]Hafez A A, Abdelaziz A Y, Hendy M A, et al. Optimal sizing of off-line microgrid via hybrid multi-objective simulated annealing particle swarm op

- imizer [J]. Computers & Electrical Engineering, 2021, 94(6): 10-19.
- [22] Saha P K, Mondal A. Efficient online heuristic approach for handling fluctuation in renewable energy in a microgrid[J]. IEEE Systems Journal, 2021, 16(2): 2471-2482.
- [23] 董清. 用并行回溯搜索算法求解 TSP 问题 [J] .黑龙江科技信息, 2011,23(33): 21-44.
- [24] Velarde P, Zafra-Cabeza A, Márquez J J, et al. Stochastic MPC-Based Reconfiguration Approaches for Microgrids [C] . IEEE Transactions on Control Systems Technology, 2023: 1-10.
- [25] Kong X, Liu X, Ma L, Lee KY. Hierarchical distributed model predictive control of standalone wind/solar/battery power system [J] . IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2019 ;49(8): 70-81.
- [26] Tudu A K, Naguru N, Dey S H N, et al. Load frequency control of an isolated microgrid using optimized model predictive control by GA[J]. Electrical Engineering, 2024, 106(1): 1-13.
- [27] Zhang Y, Meng F, Wang R, et al. Uncertainty-resistant stochastic MPC approach for optimal operation of CHP microgrid [J] . Energy, 2019, 179: 1265-1278.
- [28] 张有兵, 林一航, 黄冠弘等. 深度强化学习在微电网系统调控中的应用综述 [J] .电网技术,.2023. 47(07): 2774-2788.
- [29] Phan BC, Lee MT, Lai YC. Intelligent deep-Q-network-based energy management for an isolated microgrid. Applied Sciences [J] . 2022, 12(17): 8721.
- [30] Fang S, Zhao T, Xu Y, et al. Coordinated chance-constrained optimization of multi-energy microgrid system for balancing operation efficiency and quality-of-service [J] . Journal of Modern Power Systems and Clean Energy, 2020, 8(5): 853-862.
- [31] Gao J, Li Y, Wang B, Wu H. Multi-Microgrid Collaborative Optimization Scheduling Using an Improved Multi-Agent Soft Actor-Critic Algorithm[J]. Energies. 2023, 16(7): 32-48.
- [32] Martinelli F, Scalenghe R, Davino S ,et al. Advanced methods of plant disease detection. A review [J] . Agronomy for Sustainable Development, 201

- 5, 35(1): 1-25.
- [33]Jing Y, Zhiqiang D, Jiemai G, Siyuan C, Bing Z, Yajie W. Coordinated Control Strategy of Electricity-Heat-Gas Integrated Energy System Considering Renewable Energy Uncertainty and Multi-Agent Mixed Game [J] . Frontiers in Energy Research. 2022, 10(6): 13-23.
- [34]Huang W, Zhang C, Wu J, He X, Zhang J, Lv C. Sampling efficient deep reinforcement learning through preference-guided stochastic exploration [J] . IEEE Transactions on Neural Networks and Learning Systems. 2023,6(5): 1-12.
- [35]Yang Y, Li H, Shen B, et al. Microgrid energy management strategy base on UCB-A3C learning [J] . Frontiers in Energy Research, 2022, 10(6): 858-895.
- [36]Guo C, Wang X, Zheng Y, et al. Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning [J] . Energy, 2022, 23(8): 121-132.
- [37]孙国强, 殷岩岩, 卫志农等. 基于深度确定性策略梯度的主动配电网有功-无功协调优化调度 [J] . 电力建设, 2023, 44(11): 33-42.
- [38]Zhu Z, Lin K, Jain A K, et al. Transfer learning in deep reinforcement learning: A survey [J] . IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 44-62.
- [39]陈兴国, 孙丁源昊, 杨光等. 不动点视角下的强化学习算法综述 [J] . 计算机学报, 2023, 46(06): 46-71.
- [40]邢志伟, 张前前, 罗谦等. 基于模仿学习的机场停机位再分配决策算法 [J] . 计算机应用研究, 2022, 39(09): 65-70.
- [41]Vithayathil N, Mahmoud QH. A Survey of Multi-Task Deep Reinforcement Learning [J] . Electronics, 2020, 9(9): 13-19.
- [42]Wu J , Li H .Deep Ensemble Reinforcement Learning with Multiple Deep Deterministic Policy Gradient Algorithm [J] . Mathematical Problems in Engineering, 2020, 2020(6): 1-12.
- [43]De Bruin T, Kober J, Tuyls K, et al. Experience selection in deep reinforcement learning for control [J] . Journal of Machine Learning Research, 2018, 19(9): 1-56.

- [44]Chen D, Wang Y, Gao W. Combining a gradient-based method and an evolution strategy for multi-objective reinforcement learning [J] . Applied Intelligence. 2020, 50(10): 01-17.
- [45]Nguyen N D , Nguyen T T , Vamplew P ,et al.A Prioritized objective actor-critic method for deep reinforcement learning [J] . Neural Computing and Applications, 2021, 33(16): 10335-10349.
- [46]彭曙蓉, 彭家宜, 杨云皓等. 基于时变深度前馈神经网络的风电功率概率密度预测 [J] .电力科学与技术学报, 2023, 38(03): 84-93.
- [47]冯毅, 武卫华. 基于大数据技术的智能用电负荷预测与优化控制分析 [J] . 电子技术, 2023, 52(12): 334-335.
- [48]师洪涛, 李希彬, 丁茂生等. 基于改进 Mogrifier 门控交互机制的短期风电功率预测 [J] .中国测试, 2023, 49(08): 110-116.
- [49]周浩杰, 杨建卫, 王尊等. 基于 LSTM 光伏发电功率超短期预测模型研究 [J] . 电源技术, 2023, 47(06): 785-789.
- [50]王海军, 王磊. 一种风电集群分层模型预测控制策略研究 [J] . 电气传动, 2022, 52(11): 51-60.
- [51]宋倩芸. 大规模储能电站模块化建设方案研究 [J] .能源与环境, 2021, 27(01): 66-69.
- [52]Sioshansi R, Denholm P, Arteaga J, Awara S, Bhattacharjee S, Botterud A, Cole W, Cortes A, De Queiroz A, DeCarolis J, Ding Z. Energy-storage modeling: State-of-the-art and future research directions [J] . IEEE transactions on power systems. 2021, 37(2): 60-75.
- [53]Li C , Cao L , Chen X ,et al.Cloud Reasoning Model-based Exploration for Deep Reinforcement Learning [J] . journal of Electronics & Information Technology, 2018, 40(1): 244-248.
- [54]Icarte R T, Klassen T Q, Valenzano R, et al. Reward machines: Exploiting reward function structure in reinforcement learning [J] . Journal of Artificial Intelligence Research, 2022, 73: 173-208.
- [55]Andersen PK, Wandall EN, Pohar Perme M. Inference for transition probabilities in non-Markov multi-state models [J] . Lifetime Data Analysis. 2022 (4):

585-604.

- [56]赵娜, 张莲, 王士彬等. 并网型风光氢储微电网容量优化配置 [J]. 湖南电力, 2023, 43(04): 48-55.
- [57]刘义卿, 陈新房. 基于 CNN+LSTM 混合模型的电力消耗预测研究 [J]. 电脑与电信, 2023, 24(07): 65-69.
- [58]Kim M S, Kim J S, Choi M S, et al. Adaptive discount factor for deep reinforcement learning in continuing tasks with uncertainty [J]. Sensors, 2022, 22(19): 7266.
- [59]Cao W, Xiao J W, Cui S C, et al. An efficient and economical storage and energy sharing model for multiple multi-energy microgrids [J]. Energy, 2022, 244: 123-124.
- [60]Saha D, Bazmohammadi N, Vasquez J C, et al. Multiple microgrids: A review of architectures and operation and control strategies [J]. Energies, 2023, 16(2): 600-608.
- [61]Fu L, Meng K, Liu B, Dong ZY. Mixed-integer second-order cone programming framework for optimal scheduling of microgrids considering power flow constraints [J]. IET Renewable Power Generation. 2019, (14): 73-83.

攻读硕士期间的学术成果

- [1] Jiang Z, Lou K, Yang Z, Zhang X. Optimal Scheduling of Seasonal Microgrid based on ASAGA Algorithm. In 2023 7th International Conference on Electrical, Mechanical and Computer Engineering (ICEMCE) 2023 Oct 20 (pp. 72-77). IEEE.

尚德敬学 唯实惟新

