•新型电力系统•

DOI:10.12454/j.jsuese.202400703

本刊网刊

基于多智能体注意力–深度强化学习的输配网双侧储能协同调度方法

陈　实,朱钰杰,刘艺洪*,许刘超,唐国登

(四川大学 电气工程学院,四川 成都 610065)

摘　要:新型电力系统中储能装置受地域限制和调用手段单一因素影响,导致其利用效率低,严重制约新能源有效消纳。考虑整合输电网电源侧和配电网用户侧储能资源,通过双侧协调配合,寻求更优的储能利用方式,提升新能源消纳能力。本文提出了一种基于多智能体注意力–深度强化学习的输配网双侧储能协同调度方法。首先,考虑电源运营商的新能源消纳需求和储能供应商逐利需求,建立两者合作联盟。接着,以改进的Shapley值分配额外收入以提供合作驱动力,构建输配网双侧储能协同调度模型。进一步,将调度问题转化为马尔可夫决策过程,并提出多智能体注意力–深度强化学习算法MAAN–TD3(multi-agent attention noisy twin delayed deep deterministic policy gradient algorithm)求得最优解,该算法将注意力机制引入评估网络以捕捉智能体之间的相互依赖性,实现潜在意图识别和协同行为感知,提升算法收敛性;同时,在更新过程中注入噪声并结合软更新等机制,提升训练稳定性。以改进的IEEE输配联合系统为例进行建模求解,证明通过多智能体注意力机制能够强化合作者之间的关注度,从而兼顾到双方利益;相较于传统方法,本文方法的收敛速度和寻优效果显著提升,储能闲置时间减少7 h/d,新能源消纳量增加6.81 MW·h/d。实验结果表明,本文方法能够在有效降低联盟总成本、增加联盟各主体经济收益的同时,提高储能设备利用率、促进新能源消纳。

关键词:深度强化学习;新能源发电;共享储能;多智能体;优化调度

中图分类号:TM732　　**文献标志码:**A　　**文章编号:**2096-3246(2026)06-0148-13

随着新型电力系统建设的大力推进,新能源发电占比快速提升,但其不确定性为系统运行带来了巨大挑战。储能以其强大的灵活性支撑了新能源的消纳,提升了电力系统的调节能力[1]。然而由于储能资源分配不均、调用方式单一等问题,存在储能利用不充分的现象。充分挖掘电力系统中储能资源的灵活性,寻求多种储能的联合调度策略已经成为解决该问题的关键[2]。

近年来,共享经济蓬勃发展,共享储能是其在储能领域的延伸应用[3]。配网侧,周天娇等[4]提出储能供应商自营共享储能,考虑分时电价和负荷预测,提升了多主体经济性;南璐等[5]提出云储能双层优化模型,降低了配电网运行成本和负荷峰谷差。输电侧,Ju[6]、Song[7]等考虑新能源电站联合共享储能调度,增加了新能源消纳,达到了整体效益的最大化。上述研究虽然实现了系统的经济运行,但共享范围较为有限。在近负荷中心的输电网中,由于输配两侧电源出力特性不同,考虑储能在更大范围内的共享利用,仍有挖掘的潜力。

针对输配侧储能的协同调度已有大量相关研究提出多种储能利用方案,获得了良好的效果。Liu[8]、郝丽丽[9]等从输配全网角度出发,统筹考虑储能的配置,优化了成本并减少了新能源弃电,但未考虑多边储能共存时的具体调度方案。Dai[10]、Elliott[11]等基于输配网储能协同运行的框架,利用配电网的储能为输电网提供灵活性支撑,但未考虑其受电价的影响。Fan等[12]利用抽水蓄能电站库容和电池的剩余容量实现了输配电网的协同,促进了新能源的消纳。上述研究从全系统的视角实现了输配侧储能协同。但由于输网新能源出力特性与配网电价机制对储能要求不同,部分时段储能设备仍处于闲置状态,因此,探索更灵活高效的储能调度方式依然具有重要的研究价值。

收稿日期:2024 – 09 – 03　　**修回日期:**2024 – 11 – 15　　**网络出版日期:**2025 – 03 – 28

基金项目:国家自然科学基金项目(52377115)

作者简介:陈　实(1977—),男,副教授,博士. 研究方向:电力系统信息化及智能化、优化控制运行. E-mail:chen_shi@scu.edu.cn

* **通信作者:**刘艺洪, E-mail:1197635583@qq.com

针对多主体储能调度的复杂高维问题，传统优化方法依赖明确的目标函数，且精确建模困难，优化难度也会随着维度上升而加大[13]。作为典型的随机序贯决策问题，深度强化学习方法（DRL）结合深度神经网络和强化学习的优势，通过智能体与环境交互试错改进策略，能够较好地解决高维状态空间和动态环境的优化问题[14]。目前，DRL 算法广泛应用于配电网和微电网调度[15]、综合能源系统调度[16]等研究。陈实等[17]在配电网中应用改进的 VAE–TD3 算法，求解出多个 5G 基站储能运行的最优收益；陈实等[18]提出基于世界模型的 TD3 方法求解电力系统的低碳经济调度问题，获得了性能更优的策略；但是，上述研究所采用的单智能体 DRL 算法仅从整体考虑利益，没有考虑多主体的动态行为调整，灵活性不足，在协作环境中的学习效率和泛化性较差。Harrold 等[19]使用 MADDPG 算法控制混合储能并考虑实时价格波动，不同类型的储能均实现了效益提升；董雷等[20]基于 MADDPG 在电热联合系统进行优化，实现了电力系统和热力系统的协同合作；但是，上述算法没有区分信息的重要性，信息处理能力较差，且由于策略探索不足，收敛性较差，对全局策略的优化能力有限。

为进一步挖掘储能灵活性，本文提出基于 MAAN–TD3 的输配网双侧储能协同调度方法，以促进新能源消纳。首先，该方法在考虑各主体利益诉求的基础上构建输配网双侧储能协同模式，进一步研究输电网电源侧发电设备及储能、配电网用户侧储能的运行特性，建立输配网双侧储能协同调度模型，充分发挥储能在闲置阶段的作用，明确各利益方的经济收益。其次，提出改进强化学习算法 MAAN–TD3，在评估网络中引入注意力机制实现意图识别和协同行为感知；同时，在输入评估网络的信息中加入随机噪声扩大智能体探索空间，能确保智能体捕捉到协同者间的相互依赖性，提升学习的稳定性。最后，以合作联盟总成本为目标寻求最优运行策略，该策略能够保证有效实现电源运营商与储能供应商双侧储能协同调度，充分消纳新能源；并使用改进 Shapley 值进行收益分配，提高了联盟各主体的经济性。

1　输配网双侧储能协同模式

1.1　输配网双侧储能合作联盟的构建

在新型电力系统中，输配网双侧均建有储能设备。输电网中电源侧受储能容量限制存在新能源消纳不足的现象；配电网用户侧储能在满足用户需求的基础上通常利用剩余容量通过电价的峰谷差套利，方式较为单一，无法最大化储能效率；双方均希望有更多手段提升收益。

本文针对输配网双侧储能的现实诉求构建了配电网储能供应商与输电网电源运营商的协同模式（下文简称为储能供应商和电源运营商），双方共同组成合作联盟。储能供应商提供自身储能响应合作联盟的需求；电源运营商通过协调新能源厂站储能与用户侧储能，调动更多有效可控的灵活性资源，进一步降低联盟总成本；合作联盟在促进新能源消纳、有效提升储能利用率的同时，增加经济效益，实现双方利益共赢。合作联盟结构示意图如图 1 所示。

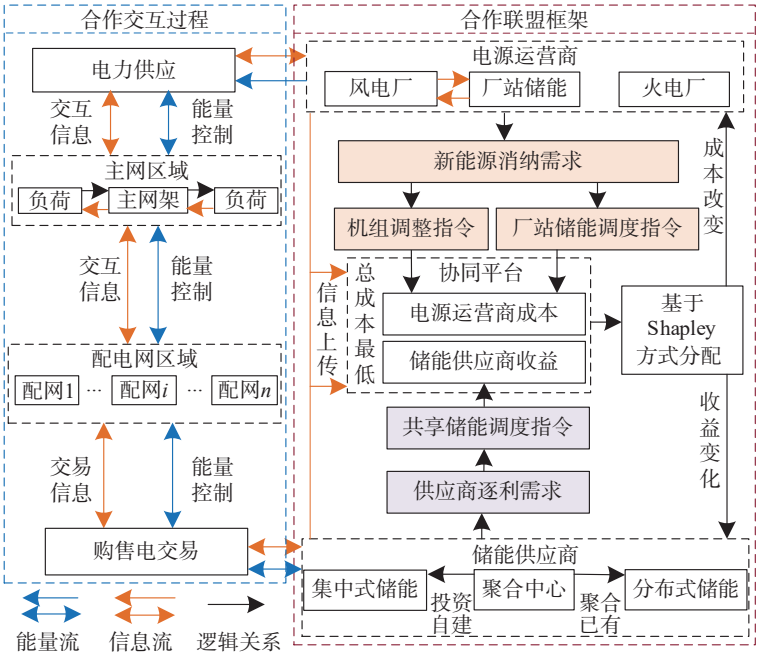


图 1　合作联盟示意图

Fig. 1　Schematic diagram of the cooperative alliance structure

图1左侧展示合作交互过程，储能运营商执行协同策略与电网交易，购电消纳电能或售电释放电能。在此交互过程的基础上，构建的合作联盟框架如图1右侧所示，电源运营商与储能供应商上传设备出力、储能荷电状态等信息至协同平台，以总成本最低求得调度策略，并以Shapley值方式分配额外收益。

1.2 合作联盟主体需求分析及协同方式设计

1.2.1 电源运营商

为应对新能源不确定性带来的电力系统安全运行风险，各地要求新能源厂站配置10%～20%的储能，但因容量有限仍无法完全消纳新能源。电源运营商希望寻求更多灵活性资源以降低不确定性带来的负面影响，提升运行性能。协同模式下，电源运营商利用厂站储能和发电设备与储能供应商开展协同调度，以实现新能源消纳。

1.2.2 储能供应商

配网侧储能供应商管理的储能设备由两部分组成：一部分是自建储能，另一部分是租赁的配网用户储能（用户出于提升用电可靠性和效率、降低用电成本等自身需求会建设分布式储能，在空闲时可以将储能出租获利）[21]。自建储能的用户依据用能情况可选择日前向储能供应商上报可租赁容量，双方签订协议，储能供应商向用户支付租赁费用。租赁期内，储能供应商拥有自建储能与用户储能容量的使用权，可通过储能控制终端发送充放电指令实现用户储能与自建储能的整体调控[22]。签订协议后，若自建储能用户出现额外的储能需求，需重新购买储能容量的使用权。

传统上，储能供应商通过两种方式获利：方式1，通过对外出租共享，面向有储能需求但因建设成本、场地等因素未配置储能的用户；方式2，使用未出租的剩余储能通过峰谷电价差套利。方式1对应的用户根据自身用电需求灵活租用储能，租用容量多变，故储能供应商可能出现容量剩余[23]。此时，储能供应商利用方式2实现峰谷差套利，但该情况下的充放电策略较单一，收益较低，储能供应商希望采取更灵活的手段追求更高的利益[3]，这正契合了输电侧电源运营商对更多灵活性资源的诉求，因此，本文研究的输配网双侧主体具备协同基础。

1.2.3 协同方式

本文提出的协同模式由协同平台实现。参与协同的电源运营商和储能供应商将可调度储能容量及新能源预测功率等共享信息上传至协同平台，平台根据合作收益最大化原则匹配合作关系。同时，电网调度中心从平台获取合作方信息，开展安全校核计算，保证系统安全。若系统能安全运行，则批准双侧储能的正常合作；否则，拒绝该协同方式，由协同平台重新匹配。

2 输配网双侧储能协同调度模型

根据上述构建的协同方式，以合作联盟运行成本最低为目标构建电源运营商和储能供应商之间的协同模型，以系统机组安全运行、电压稳定等作为约束条件。

2.1 目标函数

本文构建的协同模型以电源运营商和储能供应商的合作联盟总成本为目标函数，对应如下：

$$F=\min(F_{\text{gen}}-F_{\text{shared}}) \tag{1}$$

式中，F_{gen}为电源运营商成本，F_{shared}为储能供应商收益。

2.1.1 电源运营商成本目标

电源运营商负责集中其发电资源，生产和出售电能，保证地区的电力供应。电源运营商可能运营单一类型电厂（如火电厂）或多种类型的电厂（如风电厂、火电厂）[23]。本文设置电源运营商同时拥有火电厂和风电厂，其成本F_{gen}由火电厂成本F_{hot}和新能源电厂成本F_{renew}组成：

$$F_{\text{gen}}=F_{\text{hot}}+F_{\text{renew}} \tag{2}$$

火电厂成本主要包括机组运行成本，表示如下：

$$F_{\text{hot}}=\sum_{t=1}^{T}\sum_{j=1}^{M}(a_j(P_{j,t}^{\text{G}})^2+b_jP_{j,t}^{\text{G}}+c_j) \tag{3}$$

式中，a_j、b_j、c_j为机组j的发电成本系数，$P_{j,t}^{\text{G}}$为机组j在t时刻的出力，T为总调度时长，M为机组数量。

新能源电厂的成本目标函数F_{renew}如式(4)所示，其中，包括日常机组运维费用$F_{\text{renew-G}}$、厂站储能运维费用$F_{\text{renew-BS}}$，以及为减少弃风设置的弃风惩罚F_{quit}。

$$\begin{cases} F_{\text{renew}}=F_{\text{quit}}+F_{\text{renew-BS}}+F_{\text{renew-G}}, \\ F_{\text{quit}}=\sum_{t=1}^{T}\gamma_{\text{quit}}(-P_t^{\text{real}}+P_t^{\text{mea}}), \\ F_{\text{renew-BS}}=\sum_{t=1}^{T}\sum_{i}^{n}C_{\text{BS}}P_{i,t}^{\text{BS}}, \\ F_{\text{renew-G}}=\sum_{t=1}^{T}C_{\text{renew}}P_t^{\text{real}} \end{cases} \tag{4}$$

式中，P_t^{real}、P_t^{mea}为t时刻新能源机组的实际出力和最大可调用出力，$P_{i,t}^{\text{BS}}$为单位时间t内厂站储能i的功率，γ_{quit}为弃风惩罚系数，C_{renew}为单位功率新能源出力的成本系数，C_{BS}为厂站储能运行成本系数，n为厂站储能数量。

2.1.2 储能供应商收益目标

储能聚合商可根据电价和需求情况进行充放电。

储能供应商的收益目标函数F_{shared}如下:

$$\begin{cases} F_{\text{shared}} = F_{\text{sell}} - F_{\text{shared-loss}}, \\ F_{\text{sell}} = \sum_{t}^{T} C_t^{\text{2price}} \sum_{k}^{m} P_{k,t}^{\text{BSs}} K, \\ F_{\text{shared-loss}} = \sum_{t}^{T} \left(\sum_{k}^{m} \varphi \left| P_{k,t}^{\text{BSs}} \right| + C_t^{\text{1price}} \sum_{k}^{m} P_{k,t}^{\text{BSs}} Y \right) \end{cases} \tag{5}$$

式中:F_{sell}为售电收益;$F_{\text{shared-loss}}$为运行成本,包括购电成本和储能运维成本;φ为共享储能的单位运维成本;$P_{k,t}^{\text{BSs}}$为第k台共享储能设备t时刻的功率;C_t^{1price}为t时刻的购电电价;C_t^{2price}为t时刻的售电电价;Y与K均为0-1判别指数,购电时Y为1,K为0,售电时Y为0,K为1;m为供应商储能数量。

2.2　约束条件

协同模型的约束主要包含电网潮流、电压及储能充放电状态等。

2.2.1　电网潮流平衡约束

为保证系统稳定安全运行,需要保证电网有功功率功率平衡和无功功率平衡分别如式(6)、(7)所示:

$$P_{i,t}^{\text{G}} + P_{i,t}^{\text{renew}} + P_{i,t}^{\text{SOC}} = P_{i,t}^{\text{L}} + U_{i,t} \sum_{j \in i} U_{j,t} \left(G_{ij} \cos\theta_{ij,t} + B_{ij} \sin\theta_{ij,t} \right) \tag{6}$$

$$Q_{i,t}^{\text{G}} = Q_{i,t}^{\text{L}} + U_{i,t} \sum_{j \in i} U_{j,t} \left(G_{ij} \cos\theta_{ij,t} - B_{ij} \sin\theta_{ij,t} \right) \tag{7}$$

式(6)、(7)中,$P_{i,t}^{\text{G}}$、$P_{i,t}^{\text{renew}}$和$P_{i,t}^{\text{SOC}}$分别为火电机组、风电机组和储能t时刻注入节点i的有功功率,$Q_{i,t}^{\text{G}}$为发电机组注入节点i的无功功率,G_{ij}和B_{ij}分别为两节点间的电导和电纳,$P_{i,t}^{\text{L}}$、$Q_{i,t}^{\text{L}}$、$U_{i,t}$分别为t时刻节点i的有功功率、无功功率和电压,$\theta_{ij,t}$为t时刻线路ij上的电压相角差。

2.2.2　发电机机组功率约束

火电与风电机组为安全运行均存在极限发电功率,如式(8)所示。火电机组因机械结构限制,在相邻调度时段还受爬坡限制,产生功率上下限约束,如式(9)所示:

$$\begin{cases} P_j^{\text{G,min}} \le P_{j,t}^{\text{G}} \le P_j^{\text{G,max}}, \\ Q_j^{\text{G,min}} \le Q_{j,t}^{\text{G}} \le Q_j^{\text{G,max}}, \\ 0 \le P_t^{\text{renew}} \le P_{\text{renew,max}} \end{cases} \tag{8}$$

$$-R_{\text{dg}} \Delta t \le P_{j,t}^{\text{G}} - P_{j,t-\Delta t}^{\text{G}} \le R_{\text{dg}} \Delta t \tag{9}$$

式(8)、(9)中,$P_{j,t}^{\text{G}}$为t时刻火电机组j的出力,R_{dg}为爬坡率,$P_j^{\text{G,min}}$、$Q_j^{\text{G,min}}$、$P_j^{\text{G,max}}$、$Q_j^{\text{G,max}}$分别为火电机组j的有功最小出力、无功最小出力、有功最大出力和无功最大出力,P_t^{renew}为t时刻风电出力,$P_{\text{renew,max}}$为t时刻风电最大出力。

2.2.3　系统电压安全约束

电压安全约束表达式为:

$$U_i^{\min} \le U_{i,t} \le U_i^{\max} \tag{10}$$

式中,U_i为t时刻节点i的电压,$U_i^{\min}$和$U_i^{\max}$分别为节点i的电压上、下限。

2.2.4　联络线传输容量上下限约束

输配电网通过联络线互联,协同运行过程中需满足联络线传输容量约束,即:

$$\begin{cases} P_{\text{Trans,min}} \le P_t^{\text{Trans}} \le P_{\text{Trans,max}}, \\ Q_{\text{Trans,min}} \le Q_t^{\text{Trans}} \le Q_{\text{Trans,max}} \end{cases} \tag{11}$$

式中,P_t^{Trans}和Q_t^{Trans}分别为t时刻联络线的有功功率和无功功率,$P_{\text{Trans,min}}$、$P_{\text{Trans,max}}$和$Q_{\text{Trans,min}}$、$Q_{\text{Trans,max}}$分别为联络线有功功率的最小、最大值和联络线无功功率的最小、最大值。

2.2.5　储能充放电约束

为了储能安全运行,储能充、放电应满足容量约束,如式(12)、(13)所示;每个时段的充、放电功率也要满足极限功率约束,如式(14)、(15)所示:

$$\begin{cases} e_{t+1}^{\text{SOC}} = e_t^{\text{SOC}} + \dfrac{P_t^{\text{ch}} \eta_{\text{ch}} \Delta t}{E_{\text{cap}}}, e_{t+1}^{\text{SOC}} \le e_{\text{upper}}; \\ e_{t+1}^{\text{SOC}} = e_t^{\text{SOC}}, e_{t+1}^{\text{SOC}} > e_{\text{upper}} \end{cases} \tag{12}$$

$$\begin{cases} e_{t+1}^{\text{SOC}} = e_t^{\text{SOC}} - \dfrac{P_t^{\text{dis}} \Delta t}{\eta_{\text{dis}} E_{\text{cap}}}, e_{t+1}^{\text{SOC}} \ge e_{\text{lower}}; \\ e_{t+1}^{\text{SOC}} = e_t^{\text{SOC}}, e_{t+1}^{\text{SOC}} < e_{\text{lower}} \end{cases} \tag{13}$$

$$P_i^{\min} \le \left| P_t^{\text{ch}} \right| \le P_i^{\max} \tag{14}$$

$$P_i^{\min} \le \left| P_t^{\text{dis}} \right| \le P_i^{\max} \tag{15}$$

式(12)~(15)中,e_{t+1}^{SOC}、e_t^{SOC}和η_{ch}、η_{dis}分别为储能t+1、t时刻的SOC水平和t时刻的储能充、放电效率,e_{upper}、e_{lower}分别为储能寿命的上下限,P_t^{ch}、P_t^{dis}分别为t时刻储能充、放电功率,E_{cap}为储能的最大容量,$P_i^{\min}$、$P_i^{\max}$分别为第i台储能功率的最小值和最大值。

2.2.6　储能聚合商商业行为约束

共享储能的基本功能是保证用户日常使用,故储能聚合商应优先满足用户储能需求,之后剩余储能才可用于协同调度,故存在对应的商业行为约束[24]:

$$P_{\text{other}} t_{\text{dispatch}} \le E_{\text{allow}} - E_{\text{rent}} \tag{16}$$

式中,E_{rent}为用户的需求量,P_{other}为其他储能需求,E_{allow}为可调度的总储能容量,t_{dispatch}为调度时段。

2.3　收益分配

所构建的协同模型本质上可归结为合作博弈问题,其核心在于合理分配联盟收益。为促进新能源消纳,本文对基于Shapley值[25](具体见附录A)的传统分配方式进行了改进,引入成本贡献系数[26]校正分配权重收益,确保分配更加合理。成本贡献系数的定义为主体i的成本贡献F_i^{con}在主体成本总贡献中的占比,联盟中主体i的最终收益$f_i^{\text{'earn}}$由独立运行收益与额外

收益组成，表达式如下：

$$f_i^{'\text{earn}}=f_i^{\text{earn}}+\frac{F_i^{\text{con}}}{\sum_{i=1}^{N}F_i^{\text{con}}}\left(f_{\text{all}}^{\text{earn}}-\sum_{i=1}^{N}f_i^{\text{earn}}\right) \tag{17}$$

式中，f_i^{earn}为主体i独立运行所得收益，$f_{\text{all}}^{\text{earn}}$为联盟总收益，$N$为联盟主体数量。

储能供应商的成本贡献$F_{\text{con-s}}$表示为因参与协同产生的多余支出，具体为协同前成本$F_{\text{s-loss,n}}$与协同后成本$F_{\text{s-loss,co}}$的差距，表达式如下：

$$F_{\text{con-s}}=F_{\text{s-loss,co}}-F_{\text{s-loss,n}} \tag{18}$$

电源运营商的成本贡献$F_{\text{co-gen}}$表示为协同前弃风成本$F_{\text{quit,n}}$与协同后弃风成本$F_{\text{quit,co}}$之差，表达式为：

$$F_{\text{co-gen}}=F_{\text{quit,n}}-F_{\text{quit,co}} \tag{19}$$

3 基于 MAAN–TD3 的输配网双侧储能协同调度方法

在协同调度模式下，系统的状态维度随之上升，传统调度方法计算效率较低，策略求解困难[27]。本文将调度问题等效为马尔可夫决策过程，并基于强化学习 MAAN–TD3 算法框架求解策略。首先，在第 2 节协同模型的基础上，将调度过程转化为部分马尔可夫过程；然后，引入多头注意力机制并将其与残差结构结合以改进储能协同调度算法，在降低特征维度的同时能够识别智能体的潜在意图、感知协同行为趋势，获得更好的寻优效果；最后，提出基于协同平台的电网信息交互机制和结合软更新的算法迭代优化方法，实现策略的求解。

3.1 储能协同调度的马尔可夫决策过程

对于所研究的输配网双侧储能协同调度问题，电力系统状态随时间变化，具有随机性，且满足马尔可夫性，即下时刻状态仅依赖当前状态和动作，储能充放电可视为决策动作，则该问题可转化为部分观测马尔可夫决策过程（POMDP）模型，将决策问题形式化为状态空间$\boldsymbol{S}$、决策空间$\boldsymbol{A}$、状态转移概率$\boldsymbol{P}$和奖励函数R，并与折扣因子γ组成五元组$(\boldsymbol{S},\boldsymbol{A},\boldsymbol{P},R,\gamma)$。

1）状态空间

状态空间主要指智能体能观测到的环境的状态信息，包含共享的时间t、电网电压$\boldsymbol{V}$和功率信息$\boldsymbol{P}$，以及不同智能体部分观测的储能荷电状态$\boldsymbol{e}_{\text{SOC}}$，表示如下：

$$\boldsymbol{S}=[\boldsymbol{P},\boldsymbol{V},\boldsymbol{e}_{\text{SOC}},t] \tag{20}$$

2）动作空间

动作空间表示智能体的决策行为，即智能体对储能设备的充放电功率指令，表示如下：

$$\boldsymbol{A}=[\boldsymbol{P}_{\text{shared}},\boldsymbol{P}_{\text{gen}}] \tag{21}$$

式中，$\boldsymbol{P}_{\text{shared}}$为共享储能的充放电功率，$\boldsymbol{P}_{\text{gen}}$为厂站储能的充放电功率。

3）奖励函数

协同模型优化目标是最小化联盟总成本，求解时等价为最大化累积回报，同时，在奖励函数中引入惩罚项F_{punish}表达电压等约束，当越限程度加剧变大则影响累计回报减小，以引导设备运行在允许的范围内。智能体t时刻获得的奖励r_t表示为：

$$r_t=-F+F_{\text{punish}} \tag{22}$$

3.2 储能协同调度算法设计

3.2.1 改进思路

多智能体框架下，环境变化是非平稳过程，策略学习难度大。传统多智能体强化学习算法存在以下不足：1）输入信息是全局且高维的，缺乏特征提取，导致Q值拟合不准确；2）状态信息增多时，拼接的状态向量过长，降低了智能体处理信息的效果；3）仅对全局拟合估计，无法洞悉不同智能体的隐藏意图。为此，本文提出了 MAAN–TD3 强化学习算法，对多智能体框架下的 AC（Actor–Critic）范式进行了改进，具体思路是通过在 Critic 网络中引入注意力机制，给数据分配权重，辅助Q值拟合，以引导模型动态捕捉关键特征，达到提升全局寻优性能的目的。图 2 为其基本结构。

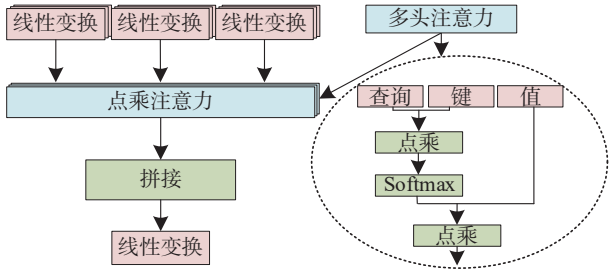


图 2 注意力机制网络结构

Fig. 2 Attention mechanism network structure

3.2.2 网络结构设计

本文将图 2 中的注意力机制改进为多头注意力后加入 Critic 网络，将状态$\boldsymbol{s}$和动作$\boldsymbol{a}$作为输入，输出评价值Q，以提升智能体的学习效率；相比于经典注意力机制，该机制能并行处理输入序列的不同部分，捕捉更丰富的特征，使智能体更易识别实现整体目标的重要特征，从而提升智能体的学习效率，对应的 Attention–Critic 结构如图 3 所示。

图 3 中，$(\boldsymbol{s},\boldsymbol{a})$表示更新过程中传输给 Critic 网络的状态动作对，网络的输出为评估值$Q(\boldsymbol{s},\boldsymbol{a})$，$\boldsymbol{X}$为原始数据，也是注意力网络的输入向量。智能体进行价值评估时，信息先经注意力机制提取特征后再得到评价值Q，综合了注意力机制优秀的特征提取能力及 AC 框架的高效性。

同时，图 3 给出了全局$(\boldsymbol{s},\boldsymbol{a})$输入评估网络后的流

程。首先，$(\boldsymbol{s},\boldsymbol{a})$经多层感知机处理，并按智能体顺序执行向量堆叠。然后，将堆叠的向量输入多头注意力网络，每个头独立提取特征后再次融合，从而甄别特征差异，捕捉智能体之间的隐藏依赖关系。输入向量经多头注意力网络得到向量$\boldsymbol{E}$，表示如下：

$$\boldsymbol{E}=\mathrm{att}((\boldsymbol{k}_1,\boldsymbol{v}_1),\boldsymbol{q}_1)\oplus\mathrm{att}((\boldsymbol{k}_2,\boldsymbol{v}_2),\boldsymbol{q}_2)\oplus\cdots\oplus\mathrm{att}((\boldsymbol{k}_m,\boldsymbol{v}_m),\boldsymbol{q}_m)\tag{23}$$

式中，$\oplus$表示将向量在同一维度拼接，att(·)代表注意力计算机制，$\boldsymbol{k}_1$、$\boldsymbol{k}_2$、$\boldsymbol{k}_m$分别为键矩阵$\boldsymbol{k}$的第1、2、m个键向量，$\boldsymbol{v}_1$、$\boldsymbol{v}_2$、$\boldsymbol{v}_m$分别为价值矩阵$\boldsymbol{v}$的第1、2、m个价值向量，$\boldsymbol{q}_1$、$\boldsymbol{q}_2$、$\boldsymbol{q}_m$分别为查询矩阵$\boldsymbol{q}$的第1、2、m个查询向量。

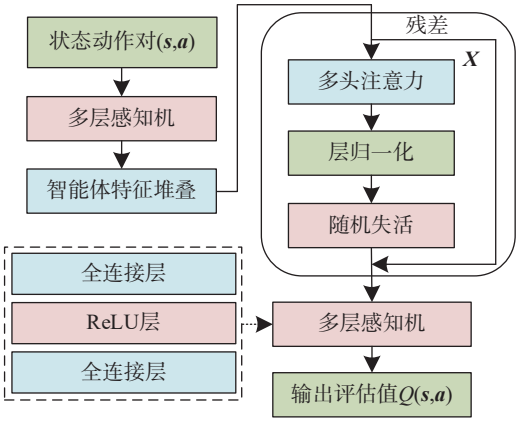


图3　Attention–Critic网络结构

Fig. 3　Attention–Critic network structure

进一步地，在注意力网络内部，将输入向量$\boldsymbol{X}$乘以参数矩阵$\boldsymbol{W}_q$、$\boldsymbol{W}_k$、$\boldsymbol{W}_v$得到对应的矩阵，具体公式为：

$$\begin{cases}\boldsymbol{q}=\boldsymbol{W}_q\boldsymbol{X},\\ \boldsymbol{k}=\boldsymbol{W}_k\boldsymbol{X},\\ \boldsymbol{v}=\boldsymbol{W}_v\boldsymbol{X}\end{cases}\tag{24}$$

在计算过程中，先将$\boldsymbol{q}$和$\boldsymbol{k}$进行点乘运算得到权重系数矩阵，该矩阵可表示$\boldsymbol{q}$和$\boldsymbol{k}$的数据相关性；再经过softmax层得到注意力分数矩阵，并与$\boldsymbol{v}$相乘后得到注意力输出，其计算如下：

$$\boldsymbol{e}=\mathrm{att}((\boldsymbol{k},\boldsymbol{v}),\boldsymbol{q})=\mathrm{softmax}\left(\frac{\boldsymbol{q}\boldsymbol{k}^{\mathrm{T}}}{\sqrt{d}}\right)\times\boldsymbol{v}\tag{25}$$

式中，d为序列的维度，$\boldsymbol{e}$为注意力机制输出向量结果。

最后，为提升训练效率，将注意力机制的输出结果经过层归一化处理；同时，引入残差网络结构ResNet，将层归一化结果与原始数据$\boldsymbol{X}$结合后再经过全连接层得到输出$\boldsymbol{X}_{\mathrm{out}}$。这里利用残差网络缓解了经典注意力机制利用方式中产生的梯度消失和特征丢失问题，且残差结构能够提升原始数据的利用率和处理效率，$\boldsymbol{X}_{\mathrm{out}}$的对应公式为：

$$\boldsymbol{X}_{\mathrm{out}}=\mathrm{FNN}(\boldsymbol{X}+\mathrm{layernorm}(\boldsymbol{e}))\tag{26}$$

式中，layernorm(·)为层归一化神经网络，FNN(·)代表前馈神经网络层。

3.3　储能协同调度方法框架及策略求解

本文提出的储能协同调度方法框架如图4所示。该框架从电网环境和算法两个层面构建，并通过协同优化机制实现输配网双侧储能的高效调度。

电网环境层面，储能协同调度策略的制定通过第1.2节所提协同平台实现。具体来说，协同双方在本地制定各自的动作策略（图4中Actor网络），并上传至协同平台；平台收集策略信息并进行价值评估（图4中Critic网络），评估值反馈至各协同方以判断策略是否需要调整，之后，将调整后的策略再次上传至平台，多次迭代并逐步优化策略，达到本地策略与全局调度需求一致的协同优化效果。

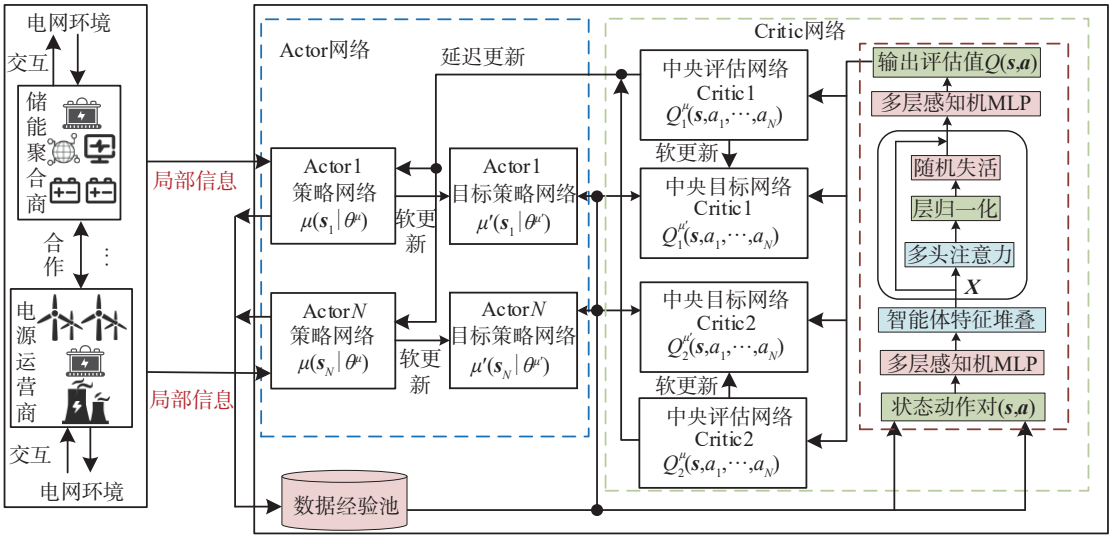


图4　基于MAAN–TD3的输配网储能调度框架

Fig. 4　Energy storage scheduling framework of transmission and distribution network based on MAAN–TD3

算法层面，MAAN-TD3 沿用传统强化学习算法 MATD3（multi-agent twin delayed deep deterministic policy gradients）的集中式训练分布式执行结构，更适用于非平稳场景，协调智能体行为，降低训练复杂度[28]。其中，集中式训练指所有智能体共用价值网络，经全连接神经网络拟合 Q 值评价动作；分布式执行指每个智能体均有独立的策略网络，由环境交互的状态输出动作。

图 4 包含了 MAAN-TD3 算法的更新过程，分为策略更新和价值更新。首先，由 Actor 策略网络与环境交互，生成动作并将交互信息存入经验池。然后，智能体从经验池提取历史数据，并在其输入 Critic 网络前注入噪声以加强探索。之后，经 Critic 网络的注意力模块生成策略评估值 Q，实现价值更新。最后，智能体依据 Q 值完成策略网络的更新，结合软更新和延迟策略更新机制提升算法训练的稳定性。其中，价值网络通过最小化估计值与 TD 误差目标的差异进行更新并提供 Q 值反馈，以指导策略网络更新，最大化累积回报，生成更优的动作。更新过程中，引入软更新和延迟策略更新机制，以避免策略和价值网络之间的冲突，噪声注入步骤及具体算法更新过程如附录 B 所示。

4 算例分析

为验证本文所提协同调度方法的有效性，选择以 IEEE30 节点和 IEEE33 节点系统相连接的改进输配联合系统进行仿真实验，33 节点系统为配网，其 1 节点与 30 节点系统（主网）的 13 节点相连接，模拟输电网靠近负荷中心（距配电网电气距离较短）的情形，因长距离输配电网的线路损耗大，协同价值有限，故本文聚焦近距离情形。风力发电厂并入主网 9 节点，容量为 100 MW，基于前述背景，风电厂按 20% 装机比例配备两台 10 MW·h/5 MW 的厂站储能；储能聚合商在配网的 40 节点和 55 节点分别配备有 3.5 MW·h/1.75 MW 的集中式储能和两台 1.5 MW·h/0.75 MW 的分布式储能，其拓扑图如图 5 所示。风力发电数据来自国内某实际系统，并乘以系数适应系统的容量，如附录 C 图 C1 所示；以某典型日负荷作为系统负荷，如附录 C 图 C2 所示。设备及运行参数分别见附录 C 表 C1 和 C2，分时电价数据见附录 C 表 C3。

以 1 h 为调度周期进行日前优化调度，训练采用 Pytorch 框架，编程环境为 Python3.7，在一套配有 Intel core i7-12700HCPU 及一张 NVIDIA RTX 3060 显卡的计算机上实现。强化学习算法训练轮次为 3 500 轮，网络权重参数更新采用 Adam 优化器。如上文模型所述，本文协同双方为储能供应商和电源运营商，对应设置两个智能体，策略网络数量 N 设置为 2，其余超参数见附录 C 表 C4 和 C5。算例设置协同与非协同两种模式对比分析，协同模式采用本文提出的储能协同调度方法，非协同模式采用第 1.2 节所述储能传统利用方式。

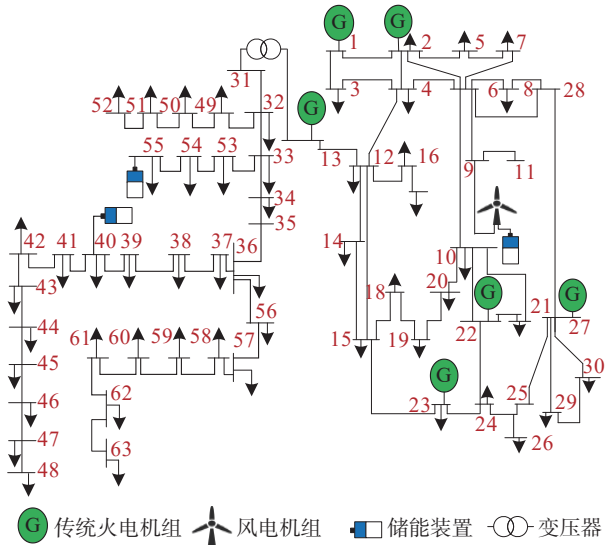


图 5 改进 IEEE 系统拓扑图

Fig. 5 Improved IEEE system topology diagram

4.1 MAAN-TD3 算法寻优性能分析

为验证本文 MAAN-TD3 算法的寻优性能，首先，将其与主流强化学习算法（MATD3 和 MADDPG）进行对比，以验证其在收敛性方面的优势；进一步，与 MAA-TD3（引入多头注意力但未注入噪声）对比验证噪声注入对算法稳定性的影响；最后，与 MACA-TD3（使用经典注意力结构）对比验证多头注意力机制加入后的效果。

图 6 展示了各算法的寻优性能对比，图 6 中阴影范围反映了给定训练轮次下奖励值在多次独立运行之间的方差。由图 6 可见：本文算法在收敛值、收敛速度和稳定性等方面均具有优势。在训练初期 500 轮次内，两种引入本文注意力机制的算法 MAAN-TD3 和 MAA-TD3 比后 3 种算法（MACA-TD3、MATD3 和 MADDPG）的奖励提升速度更快；在 400 轮次内就将奖励值从 −5.65 附近上升到 −5.25 附近，后 3 种算法在 400 轮次内奖励值仅上升到 −5.35 附近。在训练后期，约 2 000 轮次时本文算法收敛，奖励值稳定在 −5.05 附近；后 3 种算法分别在 2 000、2 500 和 3 000 轮次才趋于收敛，MATD3 和 MADDPG 的奖励值稳定在 −5.22 和 −5.25 附近，MACA-TD3 也仅稳定在 −5.09 附近，后 3 种算法在收敛速度和收敛值方面均比本文算法稍逊一筹。

本文算法性能的提升主要是因为引入的注意力机制能够确保智能体更好地提取全局特征。由图 6 中局部放大图可见：在训练后 500 轮次，本文算法的多头注意力比经典注意力机制给出的注意力结果更加全面，并且，残差网络将注意力提取到的信息与原始

信息结合，能保证信息的质量和利用效率，因此得到了更高的奖励值和更平稳的曲线；相较于 MAA-TD3，本文算法在后 500 轮次的波动更小，这是由于噪声的引入增加了策略评价网络输入的随机性，可以防止网络的过度拟合，提高稳定性。

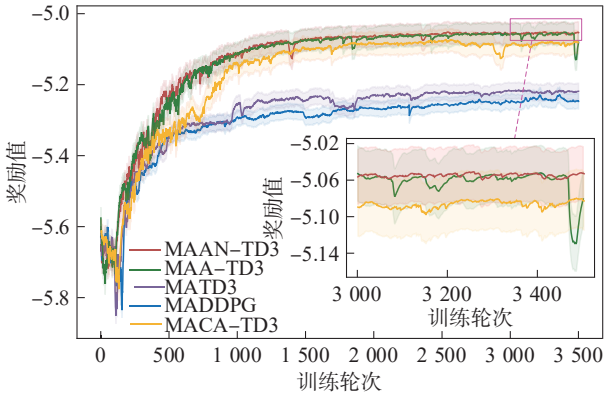


图 6　各算法寻优性能对比

Fig. 6　Optimization effect comparison of different algorithm

为验证注意力机制在训练中起到的具体作用，分别在第 1 节所述的协同模式与传统非协同模式下使用 MAAN-TD3 算法训练模型，其注意力关注度结果如图 7 所示。图 7 中，坐标轴分别表示两个智能体的自注意力和互注意力关注度，如“1→2”表示智能体 1 对智能体 2 的关注度。

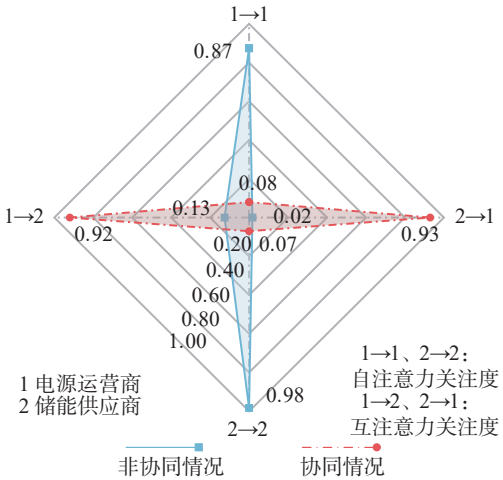


图 7　协同与非协同模式下注意力关注度对比

Fig. 7　Comparison of attention and attention in cooperative and non-cooperative mode

由图 7 可见：在非协同模式下，由于每个主体的寻优目标不同，所以更多地关注自己的信息而并不关注其他智能体信息。两个智能体的自注意力分别达到了 0.87 和 0.98，其互注意力关注度仅有 0.13 和 0.02。而协同模式下，智能体高度关注协同方，互注意力分别达到了 0.93 和 0.92。其原因是协同状态下两个智能体的利益目标相同，任一智能体的设备状态（SOC、功率约束等）和动作决策信息（发电机出力、储能充放电等）对协同方都至关重要，高度关注对方信息有助于判断潜在意图，实现动作协同，完成优化目标。总之，注意力机制的引入实质上保证了对协同关系的高关注度，以便捕捉到智能体之间的相互依赖性，这对强化学习算法的成功至关重要。

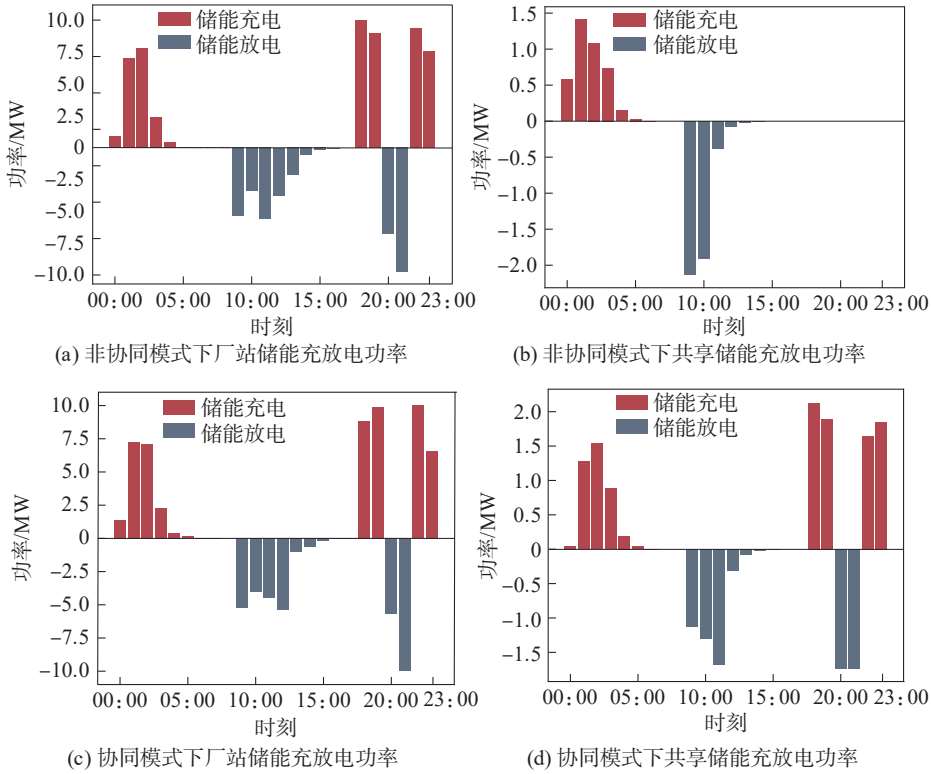
4.2　储能协同调度方法结果分析

4.2.1　储能充放电动作协同效果分析

储能协同调度方法实施前后的协同效果如图 8 所示。图 8(a)、(b)中，非协同模式下，共享储能和厂站储能根据自身目标独立开展决策。由图 8(a)、(b)可见：共享储能充放电行为跟随电价变化，在低电价时段（00:00—08:00）充电，在峰值电价区间（09:00—14:00）放电套利；厂站储能则跟随负荷变化，负荷低谷充电消纳风电，如 00:00—08:00、18:00—19:00 及 22:00—23:00，负荷增加时放电以减少火电出力降低成本，如 09:00—15:00 及 20:00—21:00。因此，非协同模式下，双侧储能的充放电选择完全出于自身目标独立决策。

协同模式下共享储能和厂站储能的充放电曲线如图 8(c)、(d)所示。由图 8(c)、(d)可以看出：在合作联盟的经济目标引导下，共享储能和厂站储能目标趋同，做出了相似的动作，实现了策略协同。具体地，在 00:00—08:00 时段，双侧储能均选择充电以减少弃风情况，充电至 SOC 最高值附近。在 09:00—23:00 时段，双侧储能展现出行为上的协同配合。其中，在 09:00—17:00 时段，负荷处于较高水平，厂站储能选择放电代替火电出力；共享储能也放电以减轻火电压力，同时兼顾电价影响，在峰值电价时段快速放电套利，SOC 降低至最低水平。在 18:00—19:00 时段，负荷降低、新能源出力增加，双侧储能同时充电消纳新能源，并准备应对晚高峰。20:00—21:00 时，负荷进入晚高峰，共享储能和厂站储能选择放电应对负荷高峰，共享储能兼顾高电价套利。22:00—23:00 时，负荷降低且新能源出力增加，双侧储能均充电以减少弃风。

图 9 为储能协同调度方法实施前后部分储能的 SOC 变化曲线。由图 9 可以看出，在协同模式下共享储能 1 d 内经历了两轮充放电，利用效率翻倍，储能闲置时间从 11 h 下降到 4 h，减少了大量闲置时间。主要表现在 18:00—21:00 时共享储能受协同目标引导充放电，在电价低的 18:00 时以最大功率充电，在电价高的 19:00 时减小充电功率，在晚高峰时段（20:00—21:00）出售全部电量。此外，第一轮的大功率放电移至 10:00—11:00 时的负荷高峰，这说明共享储能较好地跟踪到负荷，调整出力降低火电调峰压力。



(a) 非协同模式下厂站储能充放电功率　(b) 非协同模式下共享储能充放电功率

(c) 协同模式下厂站储能充放电功率　(d) 协同模式下共享储能充放电功率

图 8　协同与非协同模式下双侧储能充放电曲线

Fig. 8　Charge and discharge curves of bilateral energy storage in cooperative and non-cooperative modes

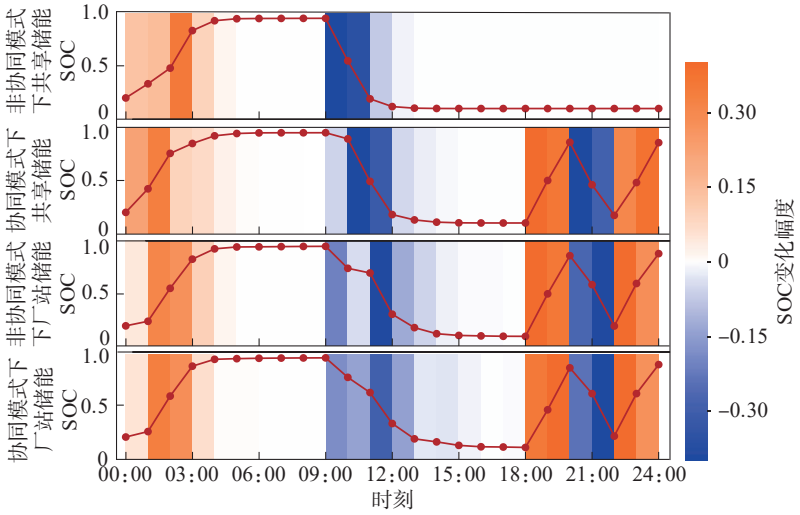


图 9　协同与非协同模式下部分储能 SOC 变化曲线

Fig. 9　Change curves of partial energy storage SOC in cooperative and non-cooperative modes

由于在总目标中电源运营商成本占比较大，厂站储能协同前后的策略变化不大，但依然可通过与共享储能的协同，进一步优化运营策略。如 09:00—11:00 负荷高峰时段，随共享储能放电时段转移，厂站储能也选择以更平缓的功率放电，体现了输配网双侧策略协同效果。

4.2.2　新能源消纳及协同收益结果分析

图 10 展示了实施储能协同调度方法后系统所有设备出力情况，图 10 中，红色曲线表示负荷的变化情况。由图 10 可见：在负荷低谷新能源出力较大的状态（如 00:00—08:00、18:00—19:00 时段等）下，调度策略能够充分利用储能的调节能力，降低火电机组出力，提升新能源利用率；负荷高峰时段，在充分调用储能的同时，火电调节机组迅速增加出力，实现功率平衡，保证负荷的用电需求，验证了本文调度方法的有效性。

储能协同调度前后发电出力数据的对比如表 1 所示。由表 1 可见，通过本文所提的输配网双侧储能，充分发挥了储能灵活性，输配网协同后 1 d 内增加风电消纳 6.81 MW·h，减少火电出力 2.53 MW·h，有效提升了新能源消纳水平。需要说明的是，本文算例通过单个配网优化调度促进新能源消纳，而实际电网中负荷中心存

在多个配网，构成一个主网与多配网相连的形态，此时运用本文调度策略具有更大潜力，新能源利用率会显著提升。

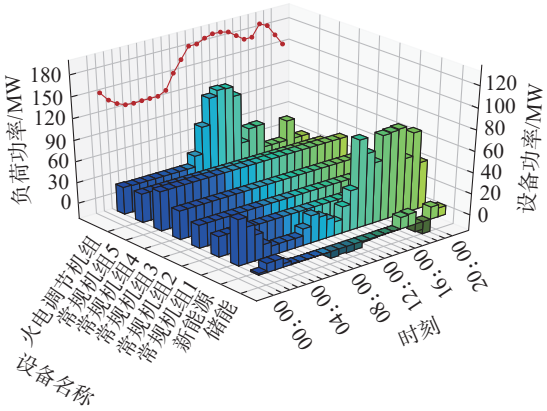


图10　协同模式下各设备出力情况

Fig. 10　Output of each device in collaborative mode

表1　协同调度前后累计发电量对比

Tab. 1　Comparison of cumulative power generation before and after coordinated scheduling

调度阶段	风电累计发电量/(MW·h)	火电调节机组累计发电量/(MW·h)
调度后	930.21	966.87
调度前	923.40	969.40

储能协同调度方法实施前后的收益情况如表2所示，其中，联盟总成本、弃风成本和储能为Shapley值法分配的收益。

表2　协同与非协同模式下收益情况对比

Tab. 2　Comparison of the revenue in the cooperative and non-cooperative mode

运行模式	联盟总成本/千元	储能供应商净收益/元	储能供应商运行成本/元	弃风成本/千元
协同模式	503.70	1 529	1 106	170.92
非协同模式	506.26	799	427	175.67
变化量	2.56	730	679	4.75

由表2可知：协同策略能够在1 d的调度中降低总成本（即联盟额外收益）2 560元，按Shapley值方式分配，电源运营商分得额外收益1 830元，储能供应商分得额外收益730元；协同后储能供应商的净收益（1 529元）较协同前（799元）几乎翻倍，达到了其追求经济效益的目标。值得注意的是，协同后储能供应商运行成本虽增加了679元，但通过策略协同减少了约4 750元弃风成本，本文改进的Shapley分配方式不仅补偿了上升的成本，还提升了供应商的净收益。电源运营商所获额外收益虽然相对其成本经济效益不大，但其减少弃风的主要优化目标效果很好。

实际电力系统中配电网储能设备与新能源电厂距离各不相同。为进一步考察两者之间的距离对协同作用的影响，按照储能设备与新能源电厂距离递增的顺序分别将配网设置于9节点、13节点和26节点进行仿真，结果如表3所示。由表3可见：配网侧储能设备距离风电越近，风电消纳作用越明显，协同后9节点多消纳了6.89 MW·h；随两者距离变远，风电消纳作用逐步减弱，26节点只能多消纳6.60 MW·h。进一步比较发现，两者距离与线损大小也呈正相关变化趋势，距离越远线路损耗越大，但线路损耗为1.2%～7.4%，相对仍在合理范围，说明输配协同后储能设备确实发挥了很好的风电消纳作用。另外，在多配网接入条件下，协同调度潜力将会更大。

表3　配网接入节点效用分析

Tab. 3　Utility analysis of distribution network access nodes

配电网配置节点	协同后风电消纳增量/(MW·h)	协同后线损增量/(MW·h)	协同后实际储能消纳增量/(MW·h)
9节点（近）	6.89	0.08	6.81
13节点（中）	6.81	0.15	6.66
26节点（远）	6.60	0.49	6.11

综上，上述算例展现出了本文方法中合作博弈的实质，即合作联盟各参与方通过注意力机制可以判断对方的潜在动作意图，从而制定自身的动作策略，兼顾新能源消纳和经济效益，获得了更大的价值回报，实现了协同优化调度目标。

5　结　论

本文提出了一种基于MAAN-TD3的输配网双侧储能协同调度方法。针对传统多智能体强化学习信息处理能力差和策略探索不足的问题，提出MAAN-TD3算法，将注意力机制与强化学习算法结合，实现了智能体间的潜在意图识别，进而通过输配网双侧行为协同优化了调度策略，并通过算例仿真验证了本文方法的有效性。通过不同场景下的调度结果对比可知，本文构建的输电网电源运营商与配电网储能供应商的协同模式，实现了源荷储能设备的有效协同；所提协同调度模型优化了储能的调用机制，储能闲置时间显著减少，有效提升了储能利用率，实现了储能更为充分、灵活的调用，同时，提升了新能源的消纳率和各主体的经济效益。

本文主要研究了储能供应商与电源运营商的协同模式及其效果，但未对多类型主体参与的协同方式及多机制融合下的协同调度进行深入探讨。未来的研究将重点考虑分布式发电商、综合能源运营商等多类型主体的参与，进一步细化模型；同时，引入碳交易及需求响应等机制，从而挖掘系统更大的调度潜力。

附录见本刊网络版，扫描标题旁的二维码可阅读网络全文。

参考文献：

[1] Xiao Xianyong, Zheng Zixuan. New power systems dominated by renewable energy towards the goal of emission peak & carbon neutrality: Contribution, key techniques, and challenges[J]. Advanced Engineering Sciences, 2022, 54(1): 47−59. [肖先勇, 郑子萱. “双碳”目标下新能源为主体的新型电力系统: 贡献、关键技术与挑战[J]. 工程科学与技术, 2022, 54(1): 47−59.]

[2] Kang Chongqing, Du Ershun, Guo Hongye, et al. Primary exploration of six essential factors in new power system[J]. Power System Technology, 2023, 47(5): 1741−1750. [康重庆, 杜尔顺, 郭鸿业, 等. 新型电力系统的六要素分析[J]. 电网技术, 2023, 47(5): 1741−1750.]

[3] Yan Dongxiang, Chen Yue. Review on business model and pricing mechanism for shared energy storage[J]. Automation of Electric Power Systems, 2022, 46(23): 178−191. [闫东翔, 陈玥. 共享储能商业模式和定价机制研究综述[J]. 电力系统自动化, 2022, 46(23): 178−191.]

[4] Zhou Tianjiao, Zhou Renjun, Huang Jingjie, et al. Energy trading method of energy storage aggregators under self-operating and sharing mode[J]. Electric Power Automation Equipment, 2023, 43(5): 171−176. [周天娇, 周任军, 黄婧杰, 等. 储能聚合商自营共享模式下电能交易方法[J]. 电力自动化设备, 2023, 43(5): 171−176.]

[5] Nan Lu, Liang Xunxing, He Chuan, et al. Bi-level optimal scheduling of distribution network−cloud energy storage system considering renewable energy uncertainties[J]. Advanced Engineering Sciences, 2025, 57(5): 142−155. [南璐, 梁迅行, 何川, 等. 计及新能源接入不确定性的配电网−云储能系统双层优化运行研究[J]. 工程科学与技术, 2025, 57(5): 142−155.]

[6] Ju Liwei, Bai Xiping, Li Gen, et al. Two-stage robust transaction optimization model and benefit allocation strategy for new energy power stations with shared energy storage considering green certificate and virtual energy storage mode[J]. Applied Energy, 2024, 362: 122996.

[7] Song Xiaoling, Zhang Huqing, Fan Lurong, et al. Planning shared energy storage systems for the spatio-temporal coordination of multi-site renewable energy sources on the power generation side[J]. Energy, 2023, 282: 128976.

[8] Liu Xinglei, Liu Jun, Chi Yuqi, et al. A coordinated planning and management framework for transmission and distribution systems with novel bilateral sharing energy storage model and time-phased consumption subsidy strategy[J]. Journal of Energy Storage, 2024, 95: 112377.

[9] Hao Lili, Lyu Xiaoxu, Ding Yixing, et al. Optimal configuration strategy of renewable energy−grid−customer side energy storage considering capacity sharing[J]. Proceedings of the CSEE, 2024, 44(14): 5607−5619. [郝丽丽, 吕肖旭, 丁逸行, 等. 考虑容量共享的新能源−电网−用户侧储能优化配置策略[J]. 中国电机工程学报, 2024, 44(14): 5607−5619.]

[10] Dai Xinliang, Guo Yi, Jiang Yuning, et al. Real-time coordination of integrated transmission and distribution systems: Flexibility modeling and distributed NMPC scheduling[J]. Electric Power Systems Research, 2024, 234: 110627.

[11] Elliott R T, Fernández-Blanco R, Kozdras K, et al. Sharing energy storage between transmission and distribution[J]. IEEE Transactions on Power Systems, 2019, 34(1): 152−162.

[12] Fan Guansheng, Lin Shunjiang, Feng Xiangyong, et al. Stochastic economic dispatch of integrated transmission and distribution networks using distributed approximate dynamic programming[J]. IEEE Systems Journal, 2022, 16(4): 5985−5996.

[13] Cao Di, Hu Weihao, Zhao Junbo, et al. Reinforcement learning and its applications in modern power and energy systems: A review[J]. Journal of Modern Power Systems and Clean Energy, 2020, 8(6): 1029−1042.

[14] Chen Xin, Qu Guannan, Tang Yujie, et al. Reinforcement learning for selective key applications in power systems: Recent advances and future challenges[J]. IEEE Transactions on Smart Grid, 2022, 13(4): 2935−2958.

[15] Huang Shuaibo, Chen Bei, Gao Jiangyu. Energy management strategy of electric vehicle charging station based on Markov decision process[J]. Electric Power Automation Equipment, 2022, 42(10): 92−99. [黄帅博, 陈蓓, 高降宇. 基于马尔可夫决策过程的电动汽车充电站能量管理策略[J]. 电力自动化设备, 2022, 42(10): 92−99.]

[16] Mo Jingshan, Yan Guangxian, Song Na, et al. Review of demand response-based optimal scheduling of electric and thermal integrated energy systems[J]. Advanced Engineering Sciences, 2025, 57(1): 296−307. [莫静山, 闫广贤, 宋娜, 等. 基于需求响应的电热综合能源系统优化调度研究综述[J]. 工程科学与技术, 2025, 57(1): 296−307.]

[17] Chen Shi, Guo Zhengwei, Zhou Buxiang, et al. Coordinated optimization scheduling for active distribution networks considering schedulable capacity of energy storage for 5G base stations[J]. Power System Technology, 2023, 47(12): 5225−5241. [陈实, 郭正伟, 周步祥, 等. 考虑5G基站储能可调度容量的有源配电网协同优化调度方法[J]. 电网技术, 2023, 47(12): 5225−5241.]

[18] Chen Shi, Zhu Yabin, Liu Yihong, et al. Low-carbon economic dispatch of wind-containing power systems based on world model deep reinforcement learning[J]. Power System Technology, 2024, 48(8): 3143−3154. [陈实, 朱亚斌, 刘艺洪, 等. 基于世界模型深度强化学习的含风电电力系统低碳经济调度[J]. 电网技术, 2024, 48(8): 3143−3154.]

[19] Harrold D J B, Cao Jun, Fan Zhong. Renewable energy integration and microgrid energy trading using multi-agent deep reinforcement learning[J]. Applied Energy, 2022, 318: 119151.

[20] Dong Lei, Liu Yu, Qiao Ji, et al. Optimal dispatch of combined heat and power system based on multi-agent deep reinforcement learning[J]. Power System Technology, 2021, 45(12): 4729−4738. [董雷, 刘雨, 乔骥, 等. 基于多智能体深度强化学习的电热联合系统优化运行[J]. 电网技术,

2021,45(12):4729-4738.]
[21] Kang Chongqing,Liu Jingkun,Zhang Ning.A new form of energy storage in future power system:Cloud energy storage[J].Automation of Electric Power Systems,2017,41(21):2-8.[康重庆,刘静琨,张宁.未来电力系统储能的新形态:云储能[J].电力系统自动化,2017,41(21):2-8.]
[22] Song Meng,Lin Gujing,Meng Jing,et al.Key technologies and applications of shared energy storage[J].Journal of Shanghai Jiao Tong University,2024,58(5):585-599.[宋梦,林固静,蒙璟,等.共享储能关键技术与应用[J].上海交通大学学报,2024,58(5):585-599.]
[23] Yu Yingqian,He Chuan,Liu Tianqi,et al.Research on robust pricing method of shared energy storage multi-package master-slave game distributed robust considering energy storage frequency modulation and life loss[J].Proceedings of the CSEE,2025,45(4):1450-1465.[喻颖倩,何川,刘天琪,等.计及储能调频与寿命损耗的多套餐共享储能主从博弈分布鲁棒定价方法研究[J].中国电机工程学报,2025,45(4):1450-1465.]
[24] Zhang Siyuan,Zhong Hao.Cloud energy storage capacity configuration for aggregated user energy storage and its transaction pricing[J].Southern Power System Technology,2022,16(11):9-19.[张思远,钟浩.聚合用户储能的云储能容量配置与定价[J].南方电网技术,2022,16(11):9-19.]
[25] Zhang Tao,Wang Jin,Mei Xin,et al.Optimal scheduling of multi-subject and multi-park cooperative game considering carbon-green certificate[J].Electric Power Automation Equipment,2024,44(9):129-135.[张涛,王金,梅欣,等.考虑碳-绿证的多主体多园区合作博弈优化调度[J].电力自动化设备,2024,44(9):129-135.]
[26] Liu Wenxia,Wang Lina,Zhang Shuai,et al.Cooperation game theory-based model for trading of power generation rights between former captive power plants and wind power[J].Power System Technology,2022,46(7):2647-2658.[刘文霞,王丽娜,张帅,等.基于合作博弈论的日前自备电厂与风电发电权交易模型[J].电网技术,2022,46(7):2647-2658.]
[27] Wang Ruonan,Dong Qi.Multiagent game decision-making method based on the learning mechanism[J].Chinese Journal of Engineering,2024,46(7):1251-1268.[王若男,董琦.基于学习机制的多智能体强化学习综述[J].工程科学学报,2024,46(7):1251-1268.]
[28] Lowe R,Wu Y I,Tamar A,et al.Multi-agent actor-critic for mixed cooperative-competitive environments[C]//NIPS'17:Proceedings of the 31st International Conference on Neural Information Processing Systems.Red Hook:Curran Associates Inc.,2017:6382-6393.

Two-sided Energy Storage Cooperative Scheduling Method for Transmission and Distribution Network Based on Multi-agent Attention-deep Reinforcement Learning

CHEN Shi, *ZHU Yujie*, *LIU Yihong**, *XU Liuchao*, *TANG Guodeng*

(College of Electrical Engineering, Sichuan University, Chengdu 610065, China)

Abstract:

Objective In modern electricity systems, energy storage devices encounter several challenges, primarily because of geographical constraints and the lack of flexible scheduling mechanisms. These limitations reduce storage efficiency and hinder the markedly effective integration of renewable energy. In this study, we propose a novel method to optimize the utilization of energy storage systems and enhance the integration and consumption of renewable energy within the power grid. Energy storage serves as a key technology in modern power systems and plays an essential role in mitigating fluctuations in renewable energy generation. However, despite advancements in energy storage technology, its efficiency in modern power grids remains constrained by insufficient coordination among multiple stakeholders. Specifically, the lack of effective integration between storage resources on the generation side (transmission network) and the consumption side (distribution network) prevents these resources from achieving their full potential. Energy storage on the generation side frequently remains underutilized due to inadequate coordination with the distribution side, whereas storage on the consumption side remains idle because of inflexible charging and discharging strategies.

Methods A multi-agent attention-deep reinforcement learning (MAAN-TD3) method was proposed to coordinate energy storage resources on the transmission and distribution sides of the grid. The method integrated storage resources on the generation side (transmission grid) and the user side (distribution grid) and coordinated their operations to establish a two-sided storage coordination model for transmission and distribution grids, improving the overall integration efficiency of renewable energy. Specifically, the method considered the distinct objectives of the two primary stakeholders: the power generator aimed to improve renewable energy consumption, whereas the storage provider sought to maximize profits. The proposed method generated additional benefits for both parties by incorporating these two objectives into the decision-making process. A modified Shapley value allocation mechanism was employed to encourage cooperation and raise mutual benefits. Another key innovation of the study involved addressing the limitations of traditional reinforcement learning algorithms in global optimization and convergence. Traditional reinforcement learning methods encountered difficulties due to the rapid expansion of state and action spaces, which hindered efficient algorithmic convergence at scale. Therefore, the study employed MAAN-TD3. This algorithm incorporated an attention mechanism into the value network, which enabled it to capture interdependencies among different agents and identify their potential intentions and collaborative behaviors. MAAN-TD3 improved convergence and global optimization performance by enabling agents to identify opportunities for cooperation more effectively. In addition, MAAN-TD3 incorporated a noise expansion mechanism that expanded the exploration space and further enhanced training stability. In MAAN-TD3, the attention mechanism enabled the system to focus selectively on relevant information, which reduced the dimensionality of the input feature space while preserving the comprehensiveness of the decision-making process. This mechanism helped the system assess the relative importance of different states and actions, improving decision-making and enabling agents to identify patterns and relationships within the system.

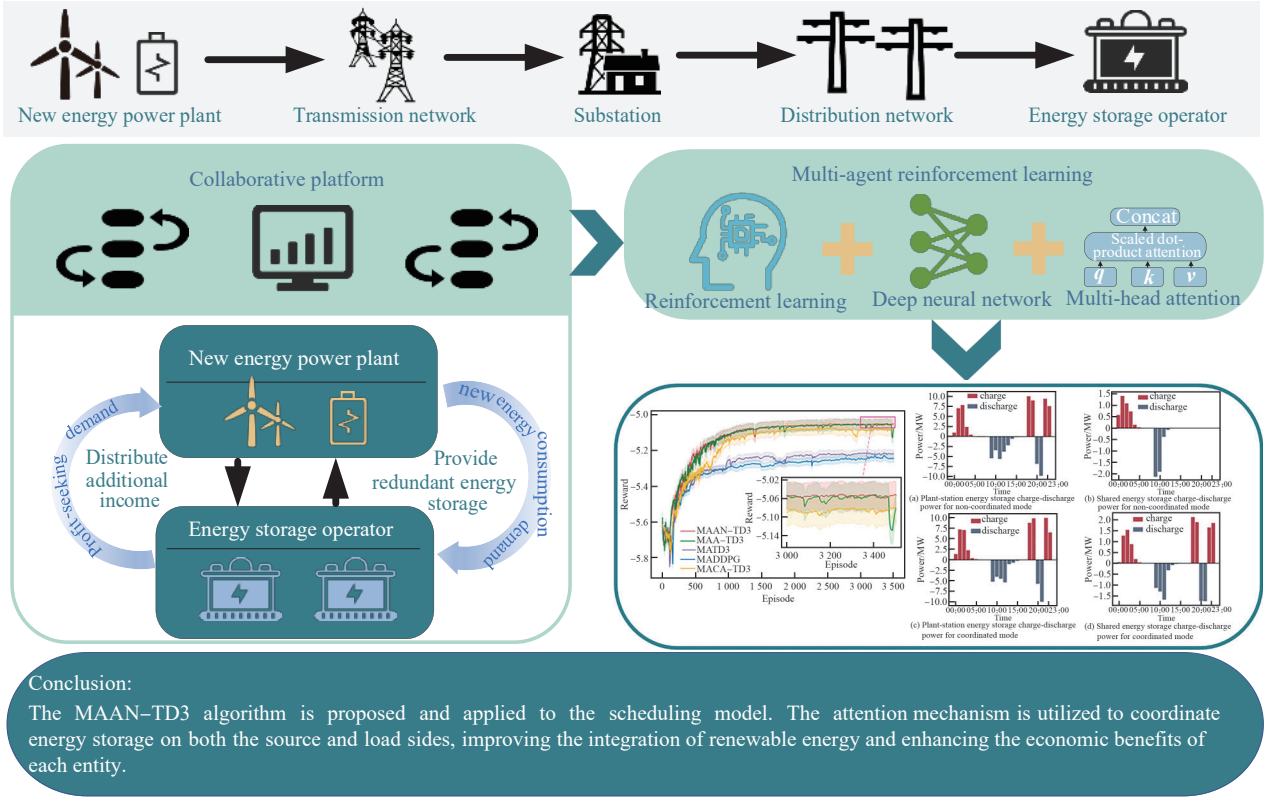
This capability was particularly important in multi-agent coordination scenarios, where system performance frequently depended on the collaborative behavior of multiple agents. The mechanism also reduced the computational burden associated with traditional algorithms when processing large-scale, high-dimensional input data. Simulation experiments were conducted using integrated IEEE 30-bus and IEEE 33-bus systems as case studies to evaluate the effectiveness of the proposed method. The simulations considered two scenarios: cooperation among stakeholders and independent operation. The study demonstrated the advantages of the proposed method by comparing and analyzing the charging and discharging strategies of energy storage devices and renewable energy consumption under these two scenarios.

Results and Discussions The simulation results demonstrated that the cooperative scheduling method based on MAAN−TD3 offered significant advantages over traditional methods. First, the proposed method optimized the charging and discharging strategies of energy storage devices and effectively reduced the total cost of the cooperative alliance. Through the coordinated optimization of storage resources on the transmission and distribution sides, storage utilization efficiency improved significantly. Second, the proposed method increased the economic returns for the power operator and storage provider, which enabled both parties to increase renewable energy consumption and improve their respective profits. Third, the proposed method enhanced the utilization of idle energy storage devices, which played a crucial role in increasing renewable energy consumption. Finally, the proposed method effectively increased overall renewable energy consumption, which represented a key objective in the green transformation of modern power systems.

Conclusions The two-sided energy storage cooperative scheduling method for transmission and distribution networks, based on multi-agent attention-deep reinforcement learning, provides an innovative and effective solution for optimizing energy storage utilization in power grids. The proposed method improves renewable energy utilization efficiency and delivers significant economic and environmental benefits to the power system by integrating energy storage resources from the transmission and distribution sides. The incorporation of the attention mechanism and noise expansion addresses the limitations of traditional methods and demonstrates superior performance in terms of convergence and global optimization. Simulation results demonstrate that the proposed method significantly reduces operating costs, enhances economic benefits, and promotes renewable energy utilization. Future studies can extend the proposed method to larger systems, further improve algorithmic performance, and investigate its application in real-world power systems, ultimately contributing to optimized energy management and sustainable energy development.

Key words: deep reinforcement learning; new energy generation; shared energy storage; multi-agent; optimal scheduling

Graphical abstract:



（编辑　赵　婧）

引用格式: Chen Shi,Zhu Yujie,Liu Yihong,et al.Two-sided energy storage cooperative scheduling method for transmission and distribution network based on multi-agent attention-deep reinforcement learning[J]. Advanced Engineering Sciences, 2026,58(6):148–160.[陈实,朱钰杰,刘艺洪,等.基于多智能体注意力−深度强化学习的输配网双侧储能协同调度方法[J].工程科学与技术,2026,58(6):148–160.]