基于深度强化学习的泛在电力物联网综合能源系统的自动发电控制

席磊^{1*}, 余璐¹, 张弦², 胡伟³

1. 三峡大学电气与新能源学院, 宜昌 443002;

2. 桂林电子科技大学电子工程与自动化学院, 桂林 541004;

3. 清华大学电机工程与应用电子技术系, 北京 100084

* E-mail: xilei2014@163.com

收稿日期: 2019-06-03; 接受日期: 2019-07-22; 网络版发表日期: 2019-12-16

国家自然科学基金(批准号: 51707102)资助项目

摘要 包含超大规模分布式能源、负荷以及冷热电联产的泛在电力物联网的综合能源系统是未来发展趋势。由于泛在物联将给电网带来强的随机扰动问题, 传统的自动发电控制(automatic generation control, AGC)方法已无法满足如此大规模复杂综合能源系统的频率稳定。机器学习是解决复杂能源系统AGC强随机扰动的一种有效方法。然而这种超大规模的泛在物联将给AGC求解带来维数灾问题。本文针对DDQN-AD(double deep Q network-action discovery)算法中经验缓存机制构建问题, 提出了一种基于比例优先级采样机制的深度强化学习算法PRDDQN-AD(prioritized replay DDQN-AD), 以解决机器学习中多维状态-动作对的维数灾问题, 进而解决泛在电力物联网综合能源系统模式下的随机扰动问题。对源网荷储协同的两区域综合能源系统模型和集成了大量源、网、荷、储及冷热电联产的多区域泛在电力物联网综合能源系统模型进行仿真。结果表明, 与改进前的DDQN-AD算法相比, PRDDQN-AD能够提升训练样本的质量, 具有良好的学习效率和泛化性能, 能够解决维数灾问题; 与其他智能算法相比, 其收敛速度和控制性能均有明显提升, 可获得区域最优协同控制。

关键词 泛在电力物联网, 综合能源系统, 深度强化学习, 自动发电控制

1 引言

十九大报告要求推进能源生产和消费革命, 构建清洁低碳、安全高效的能源体系。2019年, 国家电网有限公司在两会报告中也提出构建世界一流的能源互

联网企业。此大背景, 驱动了泛在电力物联网概念的形成。所谓泛在电力物联网, 是泛在物联网在电力行业的具体表现形式和应用落地。通过“大数据、云计算、物联网、移动互联网、人工智能”(简称“大云物移智”)尖端技术将万物互联, 表现的形式为全面感知、万物互

引用格式: 席磊, 余璐, 张弦, 等. 基于深度强化学习的泛在电力物联网综合能源系统的自动发电控制. 中国科学: 技术科学, 2020, 50: 221–234
Xi L, Yu L, Zhang X, et al. Automatic generation control of ubiquitous power Internet of Things integrated energy system based on deep reinforcement learning (in Chinese). Sci Sin Tech, 2020, 50: 221–234, doi: [10.1360/SST-2019-0199](https://doi.org/10.1360/SST-2019-0199)

联、人机交互、信息共享^[1]。而从电网频率控制的角度, 此背景下的万物互联, 将给电网带来强的随机扰动问题。传统自动发电控制(automatic generation control, AGC)方法^[2]已无法满足如此大规模复杂综合能源系统的频率稳定。

当前AGC智能控制方法可分为传统解析式和机器学习两类。基于传统解析式的控制方法, 以比例-积分-微分(proportional-integral-derivative, PID)控制方法为代表。文献[3]提出了基于社会学习自适应细菌觅食算法的最优PI/PID方法, 将社会学习机制及自适应步长策略引入到标准细菌觅食算法中, 以优化PI/PID参数, 进而提高算法收敛速度及寻优精度。文献[4]提出了一种预测优化PID方法, 以解决传统PID参数固定对系统动态特征所表现的不适应性, 从而获取最优的AGC效果。文献[5]利用分数阶PID控制器增益调整技术, 使控制器快速改变参数, 显著改善AGC的动态响应。基于机器学习的控制方法, 以强化学习(reinforcement learning, RL)的Q学习(Q-learning)算法为代表。文献[6]将Q学习算法应用于AGC中, 通过调整奖励函数中的各参数权值, 实现控制性能标准(control performance standard, CPS)的松弛控制。为加快Q学习的寻优速度, 文献[7]提出了引入资格迹的多步回溯Q(λ)算法, 有效解决了电力系统最优潮流问题。文献[8]提出了具有“赢”或“快速学习”属性的狼爬山策略RL算法, 即WoLF-PHC(λ), 通过比较当前策略与平均策略, 计算游戏的输赢, 从而快速收敛至纳什均衡。但在2×2博弈中, 对于WoLF的赢输标准, 智能体无法精确计算。因此, 文献[9]提出了具有时间隧道思想的PDWoLF-PHC(λ)算法, 各智能体利用经验共享动态更新的Q函数表, 适当调节自身控制策略, 可使多智能体系统整体学习的效率最大化。

然而, 随着分布式能源利用率的不断提高, 所带来的随机扰动也不断增加。上述方法难以满足电网的频率控制要求。因此, 为了顺应这种分布式区域调度的发展趋势^[10], 学者们提出一系列分布式的AGC控制方法, 将模型预测控制^[11]、分级递阶^[12]、自适应控制^[13]、RL^[14]以及深度学习^[15]等引入分布式AGC控制器中。然而, 随着泛在电力物联网的提出, 万物互联, 且分布式能源及负荷将继续超大规模的增加, 如风电^[16]、光伏发电^[17]、生物质发电^[18]以及冷热电联产系统^[19]等不断地大规模接入电网中, 基于泛在电力物

联网的综合能源系统也将逐渐完善。此模式下, 分布式能源及负荷的多元化与间歇性, 将给电网带来强随机扰动问题, 导致电网的频率控制难度急剧变大^[20]。因此, 需要探索新的方法以解决此模式下的强随机扰动问题。

对于低维状态-动作对的情况下, RL方法使用表格记录值函数, 为每个状态或状态-动作对分配一个存储空间, 记录其对应的函数值, 可以有效求解AGC。然而, 对于泛在电力物联网综合能源系统, 高维状态-动作对环境大规模强随机扰动所带来的维数灾问题, RL方法难以获得最优均衡解。当前, 解决维数灾问题的有效方法包括分层强化学习^[21]、迁移强化学习^[22]和神经网络值函数近似^[23]等。其中神经网络值函数近似由于具有计算高效性, 可以快速地根据输入给出输出; 模型可塑性强, 通过调整参数可改变网络模型, 具有很强的灵活性和成长性, 这些特性可以更好地解决泛在电力物联网传统RL的维数灾问题。前期, 本文作者已对基于神经网络的深度强化学习的经典算法DQN^[24]和DDQN^[25]展开研究, 开发了DQN-AD算法与具有动作探索感知思维的多智能体深度强化学习(double deep Q network-action discovery, DDQN-AD)算法^[26], 有效改善了分布式能源接入电网所带来的随机扰动问题。然而, DDQN-AD中经验回放机制主要是在记忆缓存中均匀随机地采样, 并不考虑样本的重要性, 这种方式导致学习中获取的稀缺经验数据可能被快速遗忘, 学习效率低。本文融合DDQN-AD与比例经验回放方法^[27], 提出了Prioritized Replay DDQN-AD算法, 以下简称PRDDQN-AD, 对采样的数据单元进行重要性评估, 提高经验训练效率, 以解决传统RL无法解决的由大规模泛在物联所带来的维数灾问题, 进而解决泛在电力物联网综合能源系统模式下的大规模随机扰动问题。通过大量的仿真实验, 验证了所提算法的有效性。与其他方法相比, PRDDQN-AD拥有更强的学习能力, 具有更快的收敛速度。

2 PRDDQN-AD算法

2.1 DDQN-AD算法

融合DDQN算法与AD策略, 提出了一种基于值函数迭代的DDQN-AD深度强化学习算法。通过探索环境优化更新动作集, 同时, 使用神经网络拟合RL中的

值函数 $Q(s, a; \theta) \approx Q(s, a)$, 代替了传统RL中的查表法, 解决RL的维数灾问题.

2.1.1 DDQN算法

DDQN算法采用经验回放机制, 将探索环境得到的数据以记忆单元 (s, a, r, s') 的形式作为样本储存. 随机选取样本作为当前网络的输入, 打乱数据样本的相关性, 提升控制性能. 同时DDQN拥有双神经网络 Q 和 Q^- , 具有相同的网络结构、不同的网络参数, 使用当前网络 Q 计算当前值函数, 在经过 C 步迭代计算之后, 将当前网络参数 Q 拷贝给目标网络 Q^- , 以计算目标值函数. 通过延迟参数更新的方式降低当前网络 Q 抖动, 同时减小了当前网络 Q 与目标网络 Q^- 数据的关联性. 此外, DDQN中有两套不同的参数, θ 和 θ^- . 其中, θ 用来选择对应最大 Q 值的动作, θ^- 用来评估最优动作的 Q 值. 可将动作选择和策略评估分离开, 降低了过高估计 Q 值的风险.

2.1.2 AD策略

AD策略不断地探索优化动作集合, 对于一个已知的智能体, 通过执行动作 a_i 和 a_{i+1} , 实现连续状态 s_i, s_{i+1}, s_{i+2} 的转移. 在状态 s_{i+2} , 智能体将确定是否存在一个动作 a_{i-i+2} , 可直接从状态 s_i 转移至状态 s_{i+2} , 且不属于目前动作集, 满足成本函数 $c(s_i, s_{i+2}) < \infty$. 若存在, 则智能体成功发现新动作 a_{i-i+2} , 然而将所有新探索的动作全部增添到动作集会影影响探索过程并降低学习率, 许多远离目标的动作不需要探索, 为此, 采用潜在函数 $\varphi(s)$ 对新发现的动作进行信息选择, $\varphi(s)$ 可以估计新动作状态的潜在价值, 具有精确探索目标以及降低强化学习采样复杂性的优点^[28]. 当 $\varphi(s)$ 满足可适用性时则更新动作集. AD策略打破了动作集的约束, 提高学习决策的质量, 减少获取最优策略的迭代次数.

2.2 PRDDQN-AD算法

PRDDQN-AD算法是对DDQN-AD算法的改进. DDQN-AD采用经验回放机制, 将探索环境得到的数据以记忆单元 (s, a, r, s') 的形式储存至经验池. 同时, 等概率地抽取样本用于神经网络的输入. 然而, 等概率采样并不能区分不同样本的重要性. 因此, PRDDQN-AD采用基于时间差分(temporal difference, TD)误差的

比例优先级的采样方式替代DDQN-AD中的均匀采样, 提高有价值样本的采样概率, 将每个样本的TD误差项作为评价优先级的标准. TD误差为

$$\delta = r + \gamma \max_{a'} Q(s', a'; \theta_i^-) - Q(s, a; \theta_i), \quad (1)$$

式中, r 为智能体从环境中获得的奖励函数值; γ 是折扣因子, 决定智能体的视野; θ_i 是第 i 次迭代时当前网络 Q 的参数, θ_i^- 是第 i 次迭代时目标网络 Q^- 的参数.

PRDDQN-AD的记忆单元为 $(s, a, r, s', |\delta|)$, 在算法更新的过程中, $|\delta|$ 越大, 对应记忆单元样本被采样的概率越高, 采样概率 $P(i)$ 为

$$P(i) = \frac{p_i}{\sum_{j=1}^N (p_j)}, \quad (2)$$

式中, $p_i = |\delta_i| + \sigma$, p_i 为第 i 个记忆单元中的TD误差, σ 为极小的正数, 防止 $|\delta|$ 为 0 的情况; N 为经验池容量.

同时PRDDQN-AD中引入重要性采样权重(importance-sample weights, ISW), ISW的使用放缓了参数更新的速度, 一方面导致了较慢的初始学习, 另一方面又降低了过早收敛的风险, 保证了学习的稳定性. ISW为

$$w_i = \frac{(N * P(i))^{-\beta}}{\max_j (w_j)}, \quad (3)$$

式中, w_i 为ISW, N 为经验池容量, β 为时间指数, 以权衡ISW的鲁棒性和侵略性.

损失函数为

$$L_i(\theta_i) = w_i * (r + \gamma \max_{a'} Q(s', a'; \theta_i^-) - Q(s, a; \theta_i))^2, \quad (4)$$

式中, θ_i 为第 i 次迭代时当前网络 Q 的参数.

通过神经网络的梯度反向传播更新网络参数 θ_i 和 θ_i^- :

$$\theta_{i+1} = \theta_i + \alpha * \nabla_{\theta_i} L_i(\theta_i), \quad (5)$$

$$\theta_i^- = \theta_{i+T}^-, \quad (6)$$

式中, θ_{i+1} 为第 $i+1$ 次迭代时当前网络 Q 的参数. θ_{i+T}^- 为第 $i+T$ 次迭代时当前网络 Q^- 的参数, α 为学习率.

目标 Q 值函数如式(7):

$$Y_i^{\text{PRDDQN-AD}} = \begin{cases} r_i + \gamma^{c(s_i, s_{i+1})} * Q(s_{i+1}, \max_a Q(s_{i+1}, a; \theta_i); \theta_i^-), \\ r_i, \quad s \text{ 在 } i+1 \text{ 步终止状态时,} \end{cases} \quad (7)$$

式中, r_i 为第 i 步的奖励, γ 为折扣因子。

搜索新动作并评估, 在满足成本函数小于无穷, 新探索的动作不属于原动作集, 以及其潜在函数达到可适用性这三个条件时更新动作集:

$$[c(s_i, s_{i+2}) < \infty] \wedge (a_{s_i s_{i+2}} \notin A_i(s_i)) \wedge [\gamma^{c(s_i, s_{i+2})} \varphi(s_{i+2}) > (1 + \eta) \gamma^{c(s_i, s_{i+1})} \varphi(s_{i+1})], \quad (8)$$

式中, A_i 为 i 时刻的动作集, a_{i-i+2} 为不属于当前动作集并可从状态 s_i 直接转移至状态 s_{i+2} 的动作, c 为成本函数, φ 为潜在函数, 对新发现动作进行潜在价值评估, 具有精确探索目标以及降低采样复杂性等特点, η 为松弛变量。

PRDDQN-AD 通过经验回放机制将探索环境采集的数据以记忆单元 $(s, a, r, s', |\delta|)$ 形式实时储存, 并将基

于比例优先级选取的优选样本作为当前网络的输入, 打乱数据样本的相关性, 提升样本的质量控制性能, 通过当前神经网络输出 $Q(s, a; \theta)$, 而延迟 T 步的目标神经网络输出 $\max Q(s', a'; \theta')$, 得到损失函数, 再通过随机梯度下降更新 θ 和 θ' , 从而周期性地更新目标网络参数, 得出目标 Q 值。同时在线更新动作集, 提高智能体状态转移的连通性, 减少决策次数, 加快学习速度。

3 AGC系统设计

基于PRDDQN-AD算法的泛在电力物联网综合能源系统控制架构如图1所示, 智能体全面深度感知源网荷储设备运行信息, 实时监控计算并储存互联网每

图1 (网络版彩图) 基于PRDDQN-AD的泛在电力物联网综合能源系统控制架构

Figure 1 (Color online) Ubiquitous power internet of things integrated energy system control architecture based on PRDDQN-AD.

个区域的“ACE/ Δf /CPS数据及长期历史记录”, 其中, ACE为区域控制误差(area control error, ACE), Δf 为互联电网频率偏差, CPS为控制性能指标. 计算出相应奖励值, 将当前系统环境的状态量与奖励值作为区域电网PRDDQN-AD控制器的输入, PRDDQN-AD控制器实现在线学习同时优化更新动作集, 从离散的动作集中输出最优动作, 为该区域电网调度端AGC总功率调节指令 $\Delta P_{\text{ord-}i}$, 其中, $\Delta P_{\text{T-12}}$, $\Delta P_{\text{T-11}}$, $\Delta P_{\text{T-21}}$ 为区域间的联络线交换功率.

控制器的状态与动作空间的离散化需合理设置, 如果离散化程度过细, 会使状态-动作对维数过高, 将导致决策效率低下. 反之, 又不利于改善电网频率品质. PRDDQN-AD控制器设置状态空间离散集 S 为11维, 原始动作离散集 A_0 为 $\{-50, -40, -30, -20, -10, 0, 10, 20, 30, 40, 50\}$, 同时控制器在线学习, 并探索环境优化动作集. 学习步长取AGC系统控制周期4 s.

对于泛在电力物联网综合能源系统的控制性能, 可通过CPS和频率偏差 Δf (合格范围 $\pm(0.05\sim 0.2)$ Hz)进行评估, 具体如下:

- (1) 若 $\text{CPS1} \geq 200\%$, 且 CPS2 为任意值, CPS指标合格;
- (2) 若 $100\% \leq \text{CPS1} < 200\%$, 且 $\text{CPS2} \geq 90\%$, CPS指标合格;
- (3) 若 $\text{CPS1} < 100\%$, CPS指标不合格.

3.1 奖励函数的选择

频率偏差的绝对值 $|\Delta f|$ 的大小直接影响控制效果的长期收益, 而发电成本则考虑了能量管理系统对经济性的影响. 因此, 选取第 i 次迭代中频率偏差的瞬时绝对值 $|\Delta f(i)|$ 以及第 i 次迭代中所有机组的实际发电成本的瞬时值 $C(i)_{\text{instantaneous}}$ 的加权和作为奖励函数, 加权和的绝对值越大, 将赋予智能体更小的奖励, 奖励函数如下所示:

$$R(s_i, s_{i+1}, a_i) = -\mu |\Delta f(i)|^2 - \frac{(1-\mu)C(i)_{\text{instantaneous}}}{5000}, \quad (9)$$

式中, μ 和 $1-\mu$ 分别为 $|\Delta f(i)|$ 和 $C(i)_{\text{instantaneous}}$ 奖励权重比, 这里选择 $\mu=0.5$.

3.2 参数设置

AGC控制系统的设计需要对系统参数进行合理地设置(表1).

(1) 学习率 $\alpha(0 < \alpha < 1)$. 权衡PRDDQN-AD算法的稳定性, 较大的 α 可以加快学习速度, 较小的 α 能提高系统的稳定性, 为了在保证稳定性的情况下拥有较快的学习速度, 并通过仿真试错, 选取 α 为0.5, 将会获得较好的控制效果.

(2) 折扣因子 $\gamma(0 < \gamma < 1)$. 函数的未来奖励的衰减值, 当 γ 较大时, 可以看到所有步的全部奖励, 当 γ 趋近于0时, 则只能看到当前的奖励, 由于远处的奖励最重要, 所以应该选取近似1的值. 仿真验证, 选取 γ 为0.95, 将会获得较好的控制效果.

(3) 混合策略 $\varepsilon(0 < \varepsilon < 1)$. 混合策略中按照 ε 的概率使用贪婪策略选择动作, 此外, 按照 $(1-\varepsilon)$ 的概率使用随机策略选择动作. 由于混合策略主要是选取具有最大 Q 值的动作也就是贪婪策略, 以及通过低概率的随机策略以打破局部最优, 因此选取接近于1的数值. 仿真显示, 选取 ε 为0.9, 将会获得较好的控制效果.

(4) 松弛变量 $\eta(0 < \eta < 1)$. 度量系统接受新动作的保守程度, 新动作对比于原始动作集, AD探索出来的新动作可以连接拓扑远距离状态, 提高学习效率. 为了达到最优动作集合, 通过仿真试错, 选取 η 为0.1.

(5) 时间指数 β . 时间指数 β 是一个为样本优先级引进的超参数, 用以权衡ISW中的鲁棒性和侵略性. 随着 β 的增加, ISW能够更加积极地优先考虑采样, 通过仿真发现 β 的最佳点为0.4.

(6) 经验池容量 N . 记忆单元储存在经验池中的数量, 选取 N 为100.

3.3 算法流程

PRDDQN-AD的算法流程如图2所示.

表1 AGC参数设置

Table 1 AGC parameter settings

参数	数值
学习率 α	0.5
折扣因子 γ	0.9
混合策略 ε	0.9
松弛变量 η	0.1
时间指数 β	0.4
经验池容量 N	100

图2 (网络版彩图)PRDDQN-AD算法流程

Figure 2 (Color online) PRDDQN-AD algorithm flow.

4 仿真研究

4.1 两区域综合能源系统模型

综合能源系统结构复杂, 供能方式繁多, 为了模拟分布式能源的接入对互联电网所带来的随机扰动问题, 本文在IEEE标准两区域负荷频率控制(load frequency control, LFC)模型基础上, 融入了风力发电、光伏发电、微型燃气轮机、小水电、飞轮储能和电动汽车, 建立两区域综合能源系统LFC模型。风力发电和光伏发电由于具有较大的随机性和不可控性, 作为随机负荷处理。选取调节速度较快且较为经济的微型燃气轮机与小水电作为主调频机组, 飞轮储能起辅助调频作用, 电动汽车在充放电过程中既是负荷又可以充当电源, 因此等效为一个分布式电源。在控制器运行过程中采集A和B区域的实时数据, 观测各区域当前状态以确定动作, 两个区域通过控制器数据共享和交换信息, 以期达到最优协同。在线运行之前, PRDDQN-AD需进行充分的预学习, 即通过随机探索和试错训练优化算法。

两区域综合能源系统模型如图3所示。其中, ΔP_{tie}

为联络线交换功率, T_g 为调速器时间常数, T_t 为涡轮机组时间常数, T_p 为频率响应等效函数时间常数, K_p 为频率响应等效函数系数, T_{12} 为联络线时间常数, k_{EV} 为电动汽车特性系数, α_{EV} 为电动汽车的积分比例系数, T_{EV} 为电动汽车充放电时间参数。

4.1.1 预学习正弦负荷扰动

在预学习阶段, 对A和B两区域施加时长为10000 s, 幅值 ± 1000 MW的连续正弦负荷扰动, 以10 min CPS1的平均值和频率输出为评判标准, 观察PRDDQN-AD的控制性能。图4给出了PRDDQN-AD控制器负荷扰动和A和B区域的功率输出结果, 图4(a)为预学习期间, PRDDQN-AD控制器的功率输出, 能够很好地跟随负荷扰动变化。在图4(b)和(c)中, A和B区域的CPS1(10 min的CPS1平均值)考核指标值均保持在199.5%以上, Δf_{max} (最大频率偏差)为0.012 Hz, 两个区域都具有合格的CPS和 Δf , 这说明PRDDQN-AD控制器已逼近一个确定性最优CPS控制策略。在两区域控制器均完成足够迭代次数的预学习后, 可将PRDDQN-AD控制器投入真实环境运行。

图 3 (网络版彩图)两区域综合能源系统模型

Figure 3 (Color online) Two region integrated energy system model.

图 4 (网络版彩图)PRDDQN-AD控制器的预学习过程. (a) 负荷扰动和控制功率输出曲线; (b) 10 min CPS1 平均值曲线; (c) 频率变化曲线

Figure 4 (Color online) Pre-learning process of the PRDDQN-AD controller. (a) Load disturbance and controller power output curve; (b) 10 min CPS1 average curve; (c) frequency curve.

4.1.2 阶跃负荷扰动

对两区域综合能源系统模型施加阶跃负荷扰动, 模拟大规模分布式能源接入电网所带来的负荷突增情况. 引入了仿真时长为40000 s, 幅值1000 MW的阶跃负荷扰动, 对基于PRDDQN-AD, DDQN-AD, DQN-AD, DDQN, DQN, PDWoLF-PHC(λ), WoLF-PHC(λ), Q(λ)和Q等9种算法的控制器进行仿真对比分析. 表2为9种算法的控制性能, 相较于其他算法, 所提算法的超调量降低了15.014%~77.111%, $|\Delta f|$ (频率偏差绝对值的平均值)降低了7.692%~94.906%, CPS1平均值增加了0.261%~36.326%, 因此, PRDDQN-AD具有最优的控制性能.

4.2 泛在电力物联网综合能源系统模型

泛在电力物联网是具有分布式特性的源网荷储万物互联系统, 最终形成一种数据共享的大型综合能源系统. 然而, 目前泛在电力物联网综合能源系统模型并无统一的标准. 因此, 本文在大规模分布式源网荷储互联的基础上, 融入了冷热电联产(combined cooling heating and power, CCHP)系统, 实现多种能源互补协同优化运行. 其中, CCHP系统如图5所示, 包括燃气锅炉、热交换设备、离心式制冷机, 在能量梯级利用的概念基础上, 实现了制冷、供热及发电过程一体化. CCHP系统目的在于提高能源利用效率, 减少化物及有害气体的排放, 可以大大提高能源利用效率.

泛在电力物联网综合能源系统模型包含多种中小型分布式能源, 如小水电、风电、生物质能、柴油-

图5 (网络版彩图)冷热电联产系统原理
Figure 5 (Color online) Principle of the CCHP system.

风、微燃-光伏、燃料电池-光伏等组合类型. 同时利用燃气轮机的余热, 在生产电力的同时, 满足供热、制冷两方面的需求, 如图6所示. 此外, 假定光伏、风电和电动汽车不参与调频, 只是以扰动负荷的形式存在, 故其模型进行相应简化, 建立相应的光伏出力模型; 对于风电模型, 采用有限带宽白噪声模拟的随机风作为输入, 则切入风速为3 m/s, 切除风速为20 m/s. 其中, 模型中传递函数的参数如表3所示. 各机组的相关参数如表4所示.

4.2.1 白噪声负荷扰动

为考虑更加实际的运行情况, 进一步验证所提算法的控制性能. 引入了白噪声扰动, 模拟接入未知的分布式能源的随机负荷扰动情况, 以40000 s的负荷扰动作为考核周期, 测试PRDDQN-AD, DDQN-AD, DQN-AD, DDQN, DQN, PDWoLF-PHC(λ), WoLF-PHC(λ), Q(λ)和Q 9种算法的性能. 测试结果如图7. 图7(a)显示PRDDQN-AD曲线更为平滑, 功率输出更加接近负荷扰动, 图7(b)中9种算法的 $|\Delta f|$ 平均值分别为0.0022, 0.0081, 0.0089, 0.0112, 0.0979, 0.0123, 0.0412, 0.0157, 0.0271 Hz, 显然PRDDQN-AD控制性能更优.

4.2.2 方波负荷扰动

引入方波负荷扰动, 模拟电力系统中负荷规律性突增和突减情况. 以20000 s的负荷扰动作为考核周期, 测试以上9种智能算法性能, 以进一步分析所提方法的

表2 九种算法的控制性能

Table 2 Control performance of nine algorithms

算法	超调量 (%)	$ \Delta f $ (Hz)	CPS1 (%)
PRDDQN-AD	0.1483	0.0036	199.9427
DDQN-AD	0.1851	0.0039	199.4200
DQN-AD	0.1893	0.0075	197.9616
DDQN	0.1847	0.0193	196.1595
DQN	0.1748	0.0199	196.0075
PDWoLF-PHC(λ)	0.1745	0.0233	194.8472
WoLF-PHC(λ)	0.2496	0.0318	184.2838
Q(λ)	0.4819	0.0699	153.6332
Q	0.6479	0.07067	127.3121

图 6 (网络版彩图) 泛在电力物联网综合能源系统模型

Figure 6 (Color online) Ubiquitous power internet of things integrated energy system model.

表 3 模型传递函数的参数

Table 3 Parameters of the model transfer function

机组	参数	数值
水电机组	二次时延 T_{SH}	3 s
	伺机电动机时间常数 T_P	0.04 s
	伺机增益 K_S	5
	永态转差系数 R_P	1 s
	复位时间 T_R	0.3
	暂态转差系数 R_T	1
	闸门最大开启率 $R_{maxopen}$ (pu/s)	0.16 pu/s
	闸门最大关闭率 $R_{maxclose}$ (pu/s)	0.16 pu/s
生物发电机组	机组启动时间 T_{WH}	1 s
	二次时延 T_{SB}	10 s
	调速器的时间常数 T_{GB}	0.08 s
	蒸汽启动时间 T_{WB}	5 s
微型燃气轮机机组	机械启动时间 T_{MB}	0.3 s
	二次时延 T_{SM}	5 s
	燃油系统滞后时间常数 T_1	0.08 s
	燃油系统滞后时间常数 T_2	0.3 s
	负荷限制时间常数 T_3	3 s
	温度控制环路增益 K_T	1
燃料电池机组	负荷限制 L_{max}	1.2
	二次时延 T_{SF}	2 s
	调速器的时间常数 T_F	10.056 s
柴油发电机组	逆变器增益 K_F	9.205
	二次时延 T_{SD}	7 s
	调速器的时间常数 T_{GD}	2 s
	蒸汽启动时间 T_{WF}	1 s
	机械启动时间 T_{MD}	3 s

性能. 图8(a)为不同智能算法的功率输出结果, 由图可知, PRDDQN-AD拥有更平滑的调节指令, 可以完全跟随方波负荷扰动. 图8(b)为不同智能算法的 $|\Delta f|$ 平均值. 相较于其他智能算法, PRDDQN-AD的 $|\Delta f|$ 降低了10.85%~65.70%.

4.2.3 随机负荷扰动

考虑到泛在电力物联网综合能源系统所具有的强随机性, 引入扰动幅值小于2000 MW的随机方波扰动, 进行24 h实时仿真. 以验证PRDDQN-AD的抗扰性. 图9(a)为风电场、光伏发电、电动汽车的24 h有功功率, 图9(b)显示基于PRDDQN-AD的AGC输出的总有

功率可以精确并快速地跟踪随机扰动. 图9(c)给出了各类型AGC机组的12 h出力调节曲线. 从图中可以看到, 当负荷出现正扰动时, 调节成本较低的小水电机组和微型燃气轮机优先进行正调; 当负荷出现负扰动时, 调节成本较高的生物质能发电和柴油发电等优先进行负调. 结果验证了所提方法的有效性以及快速收敛特性.

5 总结

提出了一种新颖且高效的PRDDQN-AD算法, PRDDQN-AD利用比例优先级采样, 提高了训练样本

表 4 AGC机组相关参数

Table 4 AGC unit related parameters

区域	机组类型	机组序号	$\Delta P_{Gi}^{\max}$ (kW)	$\Delta P_{Gi}^{\min}$ (kW)	C_i (\$/h)		
					a_i	b_i	c_i
区域1	水电厂	G1	250	-250	0.0001	0.0346	8.5957
		G2	150	-150	0.0001	0.0335	8.0643
		G3	150	-150	0.0001	0.0335	8.0643
		G4	100	-100	0.0001	0.0314	7.6248
	小水电	G5	100	-100	0.0001	0.0314	7.6248
	生物发电	G6	200	-200	0.0004	0.0656	8.7657
		G7	200	-200	0.0004	0.0656	8.7657
区域2	微型燃气轮机	G8	100	-100	0.0002	0.1088	5.2164
		G9	100	-100	0.0002	0.1088	5.2164
		G10	150	-150	0.0002	0.1164	5.4976
		G11	150	-150	0.0002	0.1164	5.4976
区域3	微型燃气轮机	G12	150	-150	0.0002	0.1164	5.4976
		G13	150	-150	0.0002	0.1164	5.4976
	燃料电池	G14	150	-150	0.0003	0.1189	3.5442
		G15	150	-150	0.0003	0.1189	3.5442
区域4	柴油发电机组	G16	120	-120	0.0004	0.2348	10.9952
		G17	120	-120	0.0004	0.2348	10.9952
		G18	120	-120	0.0004	0.2348	10.9952
		G19	120	-120	0.0004	0.2348	10.9952

的质量, 具有良好的学习效率和泛化性能, 相比改进前的DDQN-AD算法, 其收敛速度和训练性能均有明显提升. PRDDQN-AD利用BP神经网络近似 Q 值函数, 同时优选样本数据经验回放, 提高处理数据的效率, 打断数据关联性, 减少方差; 双网络选择评估动作, 避免神

经网络在参数中出现震荡或发散, 增加神经网络的稳定性; 在线更新动作集, 能够减少决策次数, 加快学习速度, 具有更强的计算广度和综合决策能力.

通过多种模型不同工况下的大量仿真, 结果表明, 与其他算法相比, PRDDQN-AD表现出了较强的学习

图 7 (网络版彩图)白噪声扰动下多种算法的AGC控制性能. (a) 多种算法的负荷扰动和功率输出曲线; (b) 多种算法的频率曲线

Figure 7 (Color online) AGC control performance of various algorithms under white noise perturbation. (a) Load disturbance and power output curves of various algorithms; (b) frequency curve of various algorithms.

图 8 (网络版彩图)方波扰动下多种算法的AGC控制性能。(a) 多种算法的负荷扰动和功率输出曲线;(b) 多种算法的 $|\Delta f|$ 平均值

Figure 8 (Color online) AGC control performance of various algorithms under square wave perturbation. (a) Load disturbance and power output curves of various algorithms; (b) $|\Delta f|$ average of various algorithms.

图 9 (网络版彩图)基于PRDDQN-AD的AGC机组功率曲线。(a) 光伏、风电和电动汽车24 h有功功率;(b) AGC的总有功功率和负荷扰动;(c) 不同机组的有功功率

Figure 9 (Color online) AGC unit power curve based on PRDDQN-AD. (a) 24 h active power for photovoltaic, wind and electric vehicles; (b) total active power and load disturbance of AGC; (c) active power of different units.

能力及适应性, 可在基于多智能体的泛在电力物联网综合能源系统中, 有效获取最优协同控制, 以获得AGC最优均衡解, 解决传统RL无法解决的由于大规模

泛在物联所带来高维状态-动作对的维数灾问题, 进而解决泛在电力物联网综合能源系统模式下的大规模随机扰动问题。

参考文献

- 1 Yin S R, Ai Q, Zeng S Q, et al. Energy internet multi-energy distributed optimization research challenges and prospects (in Chinese). Power Grid Technol, 2018, 42: 1359–1369 [殷爽睿, 艾芊, 曾顺奇, 等. 能源互联网多能分布式优化研究挑战与展望. 电网技术, 2018, 42: 1359–1369]

- 2 Gao Z H, Teng X L, Zhang X B. Automatic generation control strategy under the CPS standard of interconnected power grid (in Chinese). *Autom Power Syst*, 2005, 19: 40–44 [高宗和, 滕贤亮, 张小白. 互联网CPS标准下的自动发电控制策略. *电力系统自动化*, 2005, 19: 40–44]
- 3 Xie P P, Li Y H, Liu X J, et al. Design of AGC optimal PI/PID controller for interconnected power grid based on social learning adaptive bacterial foraging algorithm (in Chinese). *Proc Chin Soc Electr Eng*, 2016, 36: 5440–5448 [谢平平, 李银红, 刘晓娟, 等. 基于社会学习自适应细菌觅食算法的互联网AGC最优PI/PID控制器设计. *中国电机工程学报*, 2016, 36: 5440–5448]
- 4 Zhao X L, Lin Z Y, Fu B, et al. Application of predictive optimization PID method in AGC with wind power system (in Chinese). *J Power Syst Autom*, 2019, 31: 16–22 [赵熙临, 林震宇, 付波, 等. 预测优化PID方法在含风电电力系统AGC中的应用. *电力系统及其自动化学报*, 2019, 31: 16–22]
- 5 Kachhwaha A, Pandey S K, Dubey A K, et al. Interconnected multi-unit two-area automatic generation control using optimal tuning of fractional order PID controller along with electrical vehicle loading. In: *IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*. Delhi, 2016. 1–5
- 6 Yu T, Zhou B. CPS self-tuning control of interconnected power grid based on reinforcement learning (in Chinese). *Power Syst Protect Control*, 2009, 37: 33–48 [余涛, 周斌. 基于强化学习的互联网CPS自校正控制. *电力系统保护与控制*, 2009, 37: 33–48]
- 7 Yu T, Hu X B, Liu J. Multi-objective optimal power flow calculation based on multi-step backtracking $Q(\lambda)$ learning algorithm (in Chinese). *J South China Univ Technol (Nat Sci Ed)*, 2010, 38: 139–145 [余涛, 胡细兵, 刘靖. 基于多步回溯 $Q(\lambda)$ 学习算法的多目标最优潮流计算. *华南理工大学学报(自然科学版)*, 2010, 38: 139–145]
- 8 Xi L, Yu T, Yang B, et al. A novel multi-agent decentralized win or learn fast policy hill-climbing with eligibility trace algorithm for smart generation control of interconnected complex power grids. *Energy Convers Manage*, 2015, 103: 82–93
- 9 Xi L, Chen J, Huang Y, et al. Smart generation control based on multi-agent reinforcement learning with the idea of the time tunnel. *Energy*, 2018, 153: 977–987
- 10 Zhu X J, Zhu J H, Zhu Z Q, et al. Architecture of distributed energy network systems (in Chinese). *Chin Sci Bull*, 2017, 32: 3659–3671 [朱晓军, 朱建华, 朱振旗, 等. 分布式能源网络系统总体架构. *科学通报*, 2017, 32: 3659–3671]
- 11 Sun B H, Tang Y, Zhong W Z, et al. Advanced frequency control strategy for large-scale wind power cluster interconnected systems based on distributed model predictive control (in Chinese). *Proc CSEE*, 2017, 37: 6291–6302 [孙皓皓, 汤涌, 仲悟之, 等. 基于分布式模型预测控制的包含大规模风电集群互联系统超前频率控制策略. *中国电机工程学报*, 2017, 37: 6291–6302]
- 12 Kong X B, Liu X J, Han M. Hierarchical distributed predictive control of wind-solar hybrid power generation system (in Chinese). *Sci Sin Inform*, 2018, 48: 1316–1332 [孔小兵, 刘向杰, 韩梅. 风光互补发电系统的分级递阶分布式预测控制. *中国科学: 信息科学*, 2018, 48: 1316–1332]
- 13 Moghadam R, Modares H. Resilient adaptive optimal control of distributed multi-agent systems using reinforcement learning. *IET Control Theor Appl*, 2018, 12: 2165–2174
- 14 Xi L, Chen J F, Huang Y H, et al. Intelligent power generation control method based on multi-agent reinforcement learning with time tunnel idea (in Chinese). *Sci Sin Tech*, 2018, 48: 441–456 [席磊, 陈建峰, 黄悦华, 等. 基于具有时间隧道思想的多智能体强化学习的智能发电控制方法. *中国科学: 技术科学*, 2018, 48: 441–456]
- 15 Yin L, Yu T, Zhang X, et al. Relaxed deep learning for real-time economic generation dispatch and control with unified time scale. *Energy*, 2018, 149: 11–23
- 16 Liu L G, Liu H X, Liu Z F, et al. Three-party asymmetric evolutionary game analysis of new energy access under the wind-fire network (in Chinese). *Sci Sin Tech*, 2015, 45: 1297–1303 [刘连光, 刘鸿熹, 刘自发, 等. 新能源接入下风火网三方非对称进化博弈分析. *中国科学: 技术科学*, 2015, 45: 1297–1303]
- 17 Ma J M, Zhang C H, Zhu S Z, et al. Hydropower's support and coordinated operation of wind power solar intermittent power (in Chinese). *Sci Sin Tech*, 2015, 45: 1089–1097 [马吉明, 张楚汉, 朱守真, 等. 水电对风能太阳能间歇性电力的支持与协调运行. *中国科学: 技术科学*, 2015, 45: 1089–1097]
- 18 Hu P L, Yu Q, Zhou W, et al. Research on coal-fired coupled biomass gasification power generation technology (in Chinese). *Boiler Manuf*, 2018, 6: 32–36 [胡鹏龙, 于强, 周昊, 等. 燃煤耦合生物质气化发电技术研究. *锅炉制造*, 2018, 6: 32–36]
- 19 Yang C, Wang X S, Zhang C, et al. Characteristics of gas turbine CCHP system coupled with solar energy and compressed air (in Chinese). *China Electr Eng J*, 2017, 37: 5350–5358 [杨承, 王旭升, 张驰, 等. 太阳能与压缩空气耦合储能的燃气轮机CCHP系统特性. *中国电机工程学报*, 2017, 37: 5350–5358]

- 20 Xi L, Chen J F, Huang Y H, et al. Intelligent power generation control based on deep reinforcement learning with action self-optimization ability (in Chinese). *Sci Sin Inform*, 2018, 48: 1430–1449 [席磊, 陈建峰, 黄悦华, 等. 基于具有动作自寻优能力的深度强化学习的智能发电控制. *中国科学: 信息科学*, 2018, 48: 1430–1449]
- 21 Zhou W J, Yu Y. Review of stratified reinforcement learning (in Chinese). *J Intell Syst*, 2017, 12: 590–594 [周文吉, 俞扬. 分层强化学习综述. *智能系统学报*, 2017, 12: 590–594]
- 22 Chen Y X, Zhang X S, Guo L X, et al. Optimal carbon-energy complex flow solution for power systems based on multi-agent migration enhanced learning algorithm (in Chinese). *High Vol Technol*, 2019, 45: 863–872 [陈艺璇, 张孝顺, 郭乐欣, 等. 基于多智能体迁移强化学习算法的电力系统最优碳-能复合流求解. *高电压技术*, 2019, 45: 863–872]
- 23 Elfving S, Uchibe E, Doya K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks*, 2018, 107: 3–11
- 24 Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. *Nature*, 2015, 518: 529–533
- 25 von Hasselt H, Guez A, Silver D. Deep reinforcement learning with double Q-learning. *Comput Sci*, 2015. arXiv: 1509.06461
- 26 Xi L, Yu L, Fu Y M, et al. Automatic power generation control based on deep learning and deepening of learning perception (in Chinese). *Chin J Electr Eng*, 2019, 39: 4150–4161 [席磊, 余璐, 付一木, 等. 基于探索感知思维深度强化学习的自动发电控制. *中国电机工程学报*, 2019, 39: 4150–4161]
- 27 Schaul T, Quan J, Antonoglou I, et al. Prioritized experience replay. In: *International Conference on Learning Representations*, 2016
- 28 Banerjee B, Kraemer L. Reinforcement learning with action discovery. In: *Proceedings of the Adaptive and Learning Agents*. 2010. 30–37

Automatic generation control of ubiquitous power Internet of Things integrated energy system based on deep reinforcement learning

XI Lei¹, YU Lu¹, ZHANG Xian² & HU Wei³

¹ College of Electrical Engineering and New Energy, Three Gorges University, Yichang 443002, China;

² School of Electrical Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China;

³ Department of Electrical Engineering, Tsinghua University, Beijing 100084, China

The integrated energy systems are developing in the direction of ubiquitous power Internet of Things (IoT). The main feature is the large-scale integration of distributed energies, loads, and cogenerations, which usually brings random disturbances to the systems, thus causing frequency stability control problems, where cannot be effectively addressed by the traditional automatic generation control methods. The recently developed machine learning approach provides potential solutions for complex systems with random disturbances. However, when this approach is applied to the ultra-large-scale ubiquitous power IoT systems, the dimensionality related problem arises, and it should be solved. In this paper, a deep reinforcement learning algorithm is developed for the frequency stability control of the ultra-large-scale ubiquitous power IoT systems with random disturbances. The developed algorithm is based on the idea of a proportional priority sampling mechanism and the prioritized replay DDQN-AD (PRDDQN-AD) strategy. In this work, both the two-region integrated energy system model and the multi-regional ubiquitous power IoT integrated energy system model are adopted in simulation and analysis; these models include a large number of sources, loads, energy-storage units, and grids. Simulation and comparison results show that the training quality of samples, learning efficiency, and generalization performance of the strategy are improved by using PRDDQN-AD. The strategy has a fast convergence speed, and thus can successfully solve the dimensionality problem.

ubiquitous power internet of things, integrated energy system, deep reinforcement learning, automatic generation control

doi: [10.1360/SST-2019-0199](https://doi.org/10.1360/SST-2019-0199)