面向多区域能源互联的多智能体协同AGC策略

席磊*, 刘宗静, 孙梦梦, 武俊男

三峡大学电气与新能源学院, 宜昌 443002

* E-mail: xilei2014@163.com

收稿日期: 2020-11-23; 接受日期: 2021-03-05; 网络版发表日期: 2021-05-08

国家自然科学基金(批准号: 51707102)资助项目

摘要 传统强化学习无法有效解决由于大规模新能源接入所带来的强随机扰动, 以至产生AGC(automatic generation control)响应速度变慢、性能变差的问题. 本文以能够解决基于Q框架的强化学习算法中状态动作对的值高估问题的双层Q学习为支点, 融入了一种能够在未知的搜索空间中快速搜索最优解的灰狼优化算法, 提出一种面向多区域能源互联的多智能体协同AGC策略, 即GWDQ(grey wolf double Q)策略, 来快速获取AGC过程中多区域协同最优解. 通过对包括了混合发电燃气轮机系统、冷热电联产等多种形式能源的两区域综合能源系统模型, 及多区域能源互联的东北电网模型进行仿真, 结果显示所提策略与传统强化学习算法相比, 拥有更强的学习能力, 收敛速度和控制性能明显改善, 能够快速获取AGC过程中多区域协同最优解.

关键词 双层Q学习, 灰狼优化, AGC, 多智能体

1 引言

面向区域能源互联网的多能协同是未来的发展趋势, 然而传统多能协同策略^[1,2]在提高新能源利用率的同时, 无法有效解决大规模新能源接入所带来的强随机扰动, 以至产生AGC响应速度变慢、性能变差的问题. 为此一些学者从更加基础的自动发电控制(automatic generation control, AGC)角度, 并引入了诸如经典控制^[3]、自适应控制^[4]、鲁棒控制^[5]等方法进行探索, 以解决随机扰动带来的问题. 然而, 上述方法不能保存过去任务的经验和知识, 缺乏自学习能力, 每次新任务执行时, 需要重新初始化, 导致寻优效率较低.

基于人工智能的强化学习^[6,7]、深度强化学习^[8,9]

可通过接收环境对动作的奖励获得学习信息并更新模型参数, 可以有效利用历史信息, 拥有较强的学习能力, 被广泛应用到AGC系统中. 如文献^[6]引入Q(λ)算法, 解决火电占优的互联电网中由于火电机组大时滞环节带来的延时回报问题. 文献^[7]提出了一种基于平均报酬模型的全过程R(λ)算法, 提高考核时间段内的10 min平均控制性能标准(control performance standard, CPS)指标合格率, 获得互联电力系统CPS最优控制方法. 文献^[8]提出一种动作探索感知思维的深度强化学习DDQN-AD算法, 将神经网络的预测机制作为强化学习的动作选择机制, 获取强随机环境下的最优控制策略, 从而解决分布式能源大规模接入电网所带来的AGC性能变差的问题. 文献^[9]提出具有自动寻优能

引用格式: 席磊, 刘宗静, 孙梦梦, 等. 面向多区域能源互联的多智能体协同AGC策略. 中国科学: 技术科学, 2021, 51: 673–683

Xi L, Liu Z J, Sun M M, et al. The multiagent cooperative AGC strategy for multiregional energy interconnection (in Chinese). Sci Sin Tech, 2021, 51: 673–683, doi: 10.1360/SST-2020-0467

力的多智能体深度强化学习DDRQN-AD算法, 获得电网最优协调控制, 解决新能源以及分布式能源大规模并网带来的强随机扰动, 以及AGC响应变慢、性能变差的问题。

而现代强化学习方法大部分是在经典Q学习^[10]方法基础上衍生出来的, 这种以Q学习为框架的方法, 容易出现状态动作对的值高估问题, 从而陷入局部优化; 在未知的具有挑战性的搜索空间中, 搜索最优解时间较长。为此, 寻求解决上述问题的方法并将其应用在面向区域能源互联网的AGC控制策略中, 以获得多能协同的最优解为本文的核心。

文献^[11]提出了双层Q学习(double Q-learning, DQL)算法, 能够解决基于Q框架算法中智能体状态动作对的值高估问题。文献^[12]提出了一种模拟狼群狩猎行为的灰狼优化(grey wolf optimizer, GWO)算法, 能够在未知的搜索空间中, 快速搜索最优解。因此, 本文以DQL为支点, 融入GWO方法, 进而形成基于GWDQ(grey wolf double Q)的多智能体策略, 解决Q学习框架下容易出现状态动作对的值高估问题的同时, 获得在未知的搜索空间中探索最优解时间最短。通过搭建融入了冷热电联产及电池-燃气轮机混合发电系统(hybrid electric gas turbine, HEGT)^[13]等多种能源的两区域综合能源系统模型, 及东北电网模型进行仿真, 验证了所提策略能够在多区域能源互联网环境下, 获得多智能体协同的最优解, 进而能够解决大规模新能源接入所带来的随机扰动, 以及AGC响应速度变慢、性能变差的问题。

2 GWDQ策略

2.1 GWO

GWO^[14]模拟灰狼群体捕食行为的特性, 对猎物的追踪、包围、捕猎等过程进行搜索优化, 并且通过狼群内部的竞争在每一次更新种群时竞选出头狼来指导整体的行进方向。相比于传统Q学习算法, GWO收敛速度更快, 求解精度更高。

GWO等级制度分为3个, 分别为 α 、 β 和 ω 。设种群大小为 N 的灰狼种群为: $X=\{x_1, x_2, x_3, \dots, x_N\}$ 。 α 狼负责作出决策, 是灰狼种群中候选解的最好值; β 狼是 α 狼的候选狼, 是第二优值。当 α 狼出现意外或者能力下降时, β 狼成为新的 α 狼作出决策。 ω 狼用来平衡狼群, 并在 α 、

β 的领导下进行位置更新。GWO算法数学模型如下:

$$A = 2b \cdot r_1 - b, \quad (1)$$

$$b = 2 - 2\frac{t}{t_{\max}}, \quad (2)$$

$$C = 2r_2, \quad (3)$$

式中 r_1 和 r_2 是 $[0, 1]$ 中的随机数; t 为当前迭代次数; $t_{\max}$ 为迭代过程中最大迭代次数; 在整个迭代过程中 b 随着迭代次数增加, 从2线性减小到0; A 为收敛因子, 在灰狼算法中负责控制灰狼种群包围圈的扩散与缩小, 在 b 线性减小过程中, A 为区间 $[-b, b]$ 的随机值; 当 $|A|>1$ 时, 灰狼分散在各区域并搜寻猎物; 当 $|A|<1$ 时, 灰狼将集中搜索某个或某些区域的猎物; C 为摆动因子, 表明灰狼可以在任何一个猎物周围的搜索空间中更新它的随机位置。

$$D = |C \cdot Q^* - Q_t(s, a)|, \quad (4)$$

$$D_\alpha = |C \cdot \max Q_t(s, a_g) - Q_t(s, a)|, \quad (5)$$

$$D_\beta = |C \cdot \max Q_t(s, a_k) - Q_t(s, a)|, \quad (6)$$

$$Q_t(s, a) = Q^* + A \cdot D, \quad (7)$$

式中 $Q_t(s, a_g)$ 、 $Q_t(s, a_k)$ 分别表示当前种群中 α 、 β 在 t 次迭代中的位置向量; $Q_t(s, a)$ 表示第 t 次迭代时灰狼的位置向量, 它描述了灰狼搜索猎物时会慢慢地接近猎物并包围它的行为; Q^* 表示猎物的位置向量; D 表示灰狼与猎物之间的距离; D_α 、 D_β 分别表示当前候选灰狼与最优两条狼之间的距离。

α 、 β 具有较强识别潜在猎物位置的能力, 因此, 在每次迭代过程中, 保留当前种群中的最好两只灰狼 α 和 β , 根据它们的位置信息来更新其他搜索智能体的位置。该行为的数学模型如下:

$$Q_t^1(s, a) = \max Q_t(s, a_g) - A_1 \cdot D_\alpha, \quad (8)$$

$$Q_t^2(s, a) = \max Q_t(s, a_k) - A_2 \cdot D_\beta, \quad (9)$$

式中 $Q_t^1(s, a)$ 表示 α 在 t 次迭代时经过一次探索后的位置信息; $Q_t^2(s, a)$ 表示 β 在 t 次迭代时经过一次探索后的位置信息。

2.2 DQL

Q^[15]学习是一种基于表格值的经典强化学习算法, 它通过奖励函数影响值函数的更新, 间接监督智能体

的每一个动作, 从而保证Q学习的收敛性. DQL^[16]在Q学习基础上, 引入双估计器产生 Q 值, 参与 Q 值更新, 解决基于传统Q学习算法的状态动作对值高估问题. 同时, DQL在训练过程中使用了经验回放机制, 存储智能体与环境交互过程中得到的历史样本, 并定期抽取部分样本训练 Q 值表, 解决基于传统Q学习算法 Q 表数据关联性的问题.

在智能体学习过程中, 贪婪动作 a_g 和次贪婪动作 a_k 分别为

$$a_g(s) = \operatorname{argmax}_{a \in A} Q(s, a), \quad (10)$$

$$a_k(s) = \operatorname{argmax}_{a \in A} Q_k(s, a), \quad (11)$$

式中 $Q(s, a)$ 为状态 s 在动作 a 下贪婪策略的 Q 值函数; $Q_k(s, a)$ 为状态 s 在动作 a 下次贪婪策略的 Q 值函数.

两个 Q 值函数误差分别为

$$\delta_n = R(s_n, s_{n+1}, a_n) + k\gamma_1 \max Q_n(s_{n+1}, a_g) - Q_n(s_n, a_n), \quad (12)$$

$$\delta_{k,n} = R(s_n, s_{n+1}, a_n) + k\gamma_1 \max Q_{k,n}(s_{n+1}, a_k) - Q_{k,n}(s_n, a_n), \quad (13)$$

式中 δ_n 为智能体在第 n 次迭代过程中的 Q 值函数误差; $\delta_{k,n}$ 为智能体在第 n 次迭代过程中的 Q_k 值函数误差; $R(s_n, s_{n+1}, a_n)$ 为智能体在选取的动作 a_n 下, 由状态 s_n 转移到状态 s_{n+1} 的奖励函数; γ_1 为奖励折扣因子; $Q_n(s_{n+1}, a_g)$ 为在贪婪动作 a_g 下, 状态 s_{n+1} 时的 Q 值函数; $Q_{k,n}(s_{n+1}, a_k)$ 为在次贪婪动作 a_k 下, 状态 s_{n+1} 时的 Q_k 值函数.

值函数学习迭代公式为

$$Q_{n+1}(s_n, a_n) = Q_n(s_n, a_n) + \psi \delta_{k,n}, \quad (14)$$

$$Q_{k,n+1}(s_n, a_n) = Q_{k,n}(s_n, a_n) + \psi \delta_n, \quad (15)$$

式中 ψ 表示值函数的学习率, 衡量GWDQ策略的稳定性.

2.3 GWDQ

GWDQ策略流程图如图1所示.

3 AGC系统设计

AGC控制系统的设计需要保证CPS^[17]、区域控制误差(area control error, ACE)^[18]和频率偏差 Δf 三个重

图1 GWDQ策略流程图

Figure 1 GWDQ algorithm flow.

要的控制性能指标在合理的范围内. Δf 的合格范围为 $\pm(0.05 \sim 0.2)$ Hz^[19]; ACE应尽可能最小化, 计算公式为

$$ACE = \Delta P_t - 10B\Delta f, \quad (16)$$

式中 B 为每个区域的频率响应系数, 单位为MW/0.1Hz.

CPS考核标准由CPS1和CPS2组成, 其中, CPS1是统计ACE变化量与频率变化关系的标准, 用于提高频率控制质量; CPS2能够限制大的不可接受且不可预见的系统潮流, 控制ACE误差在合理的范围内. CPS的考核标准具体如下:

(1) 若 $CPS1 \geq 200\%$, 则 $CPS2$ 为任意值时, CPS 指标合格;

(2) 若 $100\% \leq CPS1 < 200\%$, 且 $CPS2 \geq 90\%$, CPS 指标合格;

(3) 若 $CPS1 < 100\%$, CPS 指标不合格.

3.1 奖励函数

本文以 ACE 和 $CPS1$ 的线性加权组成奖励函数. 各区域电网的奖励函数如下所示:

$$R(s_{k-1}, s_k, a_{k-1}) = -\eta[ACE(t)]^2 - \frac{(1-\eta)[CPS1(t)]}{1000}, \quad (17)$$

式中 $ACE(t)$ 代表 t 时刻 ACE 的瞬时值, $CPS1(t)$ 为 t 时刻的 $CPS1$ 的数值, η 和 $1-\eta$ 分别表示 ACE 和 $CPS1$ 的权值. 将 ACE 和 $CPS1$ 看作相同的权重, η 取为0.5.

在AGC系统中, 智能体的输出动作为实际电网调度端的总发电调节指令, 状态为实时监测数据库以及长期历史记录数据库中当前系统的 ACE 、 Δf 、功率偏差等指标. 基于GWDQ算法的控制器分析系统特性, 需确定状态空间离散集 S 和原始动作集 A_0 . 如果离散化程度太细, 会使 Q 矩阵维数过高且导致AGC机组频繁反复发送调节指令, 反之, 又不利于改善频率品质, 因此合理安排状态与动作空间离散化十分重要. 通过机组特性以及仿真测试, 设置原始动作离散集为 $A_0 = \{-50, -45, -40, -35, -30, -25, -20, -15, -10, -5, 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50\}$. 学习步长取AGC系统控制周期, 该标准算例中取4 s.

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.

(1) 值函数学习因子 $\Psi(0 < \Psi < 1)$, 权衡GWDQ策略的稳定性, 较大的 Ψ 可以加快 Q 和 Q_k 值函数的迭代更新速度, 而较小的 Ψ 能提高系统的稳定性. 仿真显示本文控制算法的学习因子 $\Psi=0.1$ 时, 控制系统性能更优.

(2) 奖励折扣因子 $\gamma_1(0 < \gamma_1 < 1)$, 函数的未来奖励的衰减值. 在负荷频率控制过程中, 最新的奖励最为重要. 为提高算法的控制效果, γ_1 应近似于1. 通过仿真试错, γ_1 为0.9将会获得良好的控制效果.

(3) 随机数 $r(0 < r < 1)$, r_1 为 $[0, 1]$ 区间中的随机数, 确定灰狼搜索的摆动因子; r_2 为 $[0, 1]$ 区间中的随机数, 确

定灰狼搜索的收敛因子.

(4) 最大迭代数 $t_{\max}$, GWDQ终止的条件之一, 仿真显示GWDQ最大迭代次数 $t_{\max}=10$, 控制系统性能更优.

(5) 种群规模 N , 本文 $N=2$, 一个是DQL算法部分的多智能体; 另一个是GWO算法部分的灰狼种群.

(6) 灰狼控制参数 b , 能够协调算法的全局探索和局部利用能力, 并对群体多样性和收敛速度产生较大的影响. 随着迭代次数增加, b 从2线性减少到0, 用于确定灰狼搜索过程的收敛因子.

4 算例分析

4.1 改进的IEEE标准两区域负荷频率控制模型

本文在IEEE标准两区域基础上, 加入水电、风电、火电、柴油发电、光伏、生物质能、燃料电池、微型燃气轮机、飞轮储能^[20]、冷热电联产^[21], 以及HEGT, 建立了图2的两区域综合能源系统模型, 机组参数如表1所示. 其中, ΔP_{tie} 为联络线交换功率. 风电和光伏具有很大的不可控性, 因此仅作为随机负荷处理. HEGT^[22,23]的控制模型如图3所示, 包括了速度和加速度控制、燃料控制、燃气轮机、温度控制, 以及电池储能等5部分. P_{ref} 表示负荷参考值, ω_g 为发电机转速信号, Z 代表控制器的控制模式. 其中, HEGT模型中的排气温度和转矩方程为

$$f_1 = T_R - 700(1 - W_f) - 550(1 - \omega), \quad (18)$$

$$f_2 = 1.3(W_f - 0.23) + 0.5(1 - \omega), \quad (19)$$

式中 ω 为燃机转速, W_f 为燃料流量信号, T_R 为排气温度基准, 由微型燃气轮机的类型决定, 取 $T_R=950^\circ\text{C}$.

4.1.1 预学习训练

在线运行前, 需要对系统进行大量的预学习训练, 优化 Q 函数和状态动作集. 在预学习过程中, 对两区域电网输入一个连续的正弦变量(周期1200 s, 幅值1000 MW). 由图4可知多智能体的探索过程: 初始 Q 值为0, 经过大约2000 s的训练后, 探索到最优的状态动作对应的 Q 值. 图5为预学习过程中A区域总有功功率跟随正弦扰动的过程, 在初始的微弱误差后, GWDQ的输出可以准确跟踪输入变量. 由图6可以看出, 初始频率偏差在0.03 Hz之内, 经过2000 s的预学习后, 频率偏差控制在0.01 Hz以内; ACE 的10 min考核指标值保

图 2 (网络版彩图)两区域综合能源系统模型
Figure 2 (Color online) The model of two-area integrated energy system.

表 1 两区域综合能源系统模型参数设置
Table 1 The parameter settings in two-area integrated energy system model

机组	参数	数值
微型燃气轮机	T_{SM} (二次时延)	5 s
	T_1 (燃油系统滞后时间常数1)	0.08 s
	T_2 (燃油系统滞后时间常数2)	0.3 s
	T_3 (负荷限制时间常数)	3 s
	K_T (温度控制环路增益)	1
	L_{max} (负荷限制)	1.2
火电	T_g (调速器时间常数)	0.08 s
	T_s (二次延时)	15 s
	T_p (频率响应等效函数时间常数)	20 s
	T_{12} (联络线时间常数)	3.42 s

持在45 MW以下，且考核指标值在最初扰动后趋于稳定。

4.1.2 随机正弦扰动

模拟实际工况下的随机特性，以12 h的随机正弦扰动作为考核周期，测试GWDQ在实际电网中的长期性能，并截取了第一小时中各发电机组的功率情况，如图7和8。从图7可知，总机组功率可以精确并快速地跟踪负荷扰动。图8给出了12 h负荷扰动过程中，在第一个小时内AGC各个机组有功功率。当电网中出现随机性负荷扰动时，HEGT机组快速启动调节出力，以应对电网的负荷波动；水电、微型燃气轮机、生物发电

图 3 (网络版彩图) HEGT控制模型
Figure 3 (Color online) HEGT control model.

图 4 (网络版彩图) Q函数探索最优解过程
Figure 4 (Color online) Q function to explore the optimal solution process.

图 6 (网络版彩图)性能指标
Figure 6 (Color online) Control performance.

图 5 (网络版彩图)预学习过程结果
Figure 5 (Color online) The results of pre-learning process.

等清洁能源机组随后进行正调；火电、柴油发电等传统机组在扰动大时，加大发电量，在系统稳定后，逐渐减少发电量，实现对清洁能源机组高利用率的同时减少了能源的浪费。

4.1.3 阶跃扰动

对两区域引入阶跃扰动，模拟电力系统中负荷突

图 7 AGC机组的总有功功率和负荷扰动
Figure 7 Total active power and load disturbance of AGC.

图 8 不同机组有功功率
Figure 8 Active power of different units.

增情况. 对在A区域Q, DQL, PDWOLF-PHC(λ), DDRQN, DDQN-AD, GWO, PRDDQN-AD, GWDQ 8种算法进行对比分析. 由图9可知, GWDQ的CPS1更平稳, 最小值199.78%, 更接近最优值200%. GWDQ频率波动最大幅值为0.005 Hz, 相较其他算法波动最小, 更快速收敛到50 Hz.

使用一个2范数的Q函数矩阵 $\|Q_k(s,a)-Q_{k-1}(s,a)\|_2 \leq \zeta$ (ζ 为已知常量)作为预学习的终止条件, 预学习后保存Q值和look-up表. A区域在预学习过程中不同算法下的Q函数偏差收敛情况如图10所示, GWDQ波动小, 收敛快速, 相比其他算法提高了95.45%~353.81%.

4.1.4 方波扰动

对两区域施加方波扰动, 模拟电力系统中负荷规律突增和突减情况, 控制性能如图11. 图中 $|\Delta f|$ (频率偏差绝对值)为实际 $|\Delta f|$ 扩大1000倍的平均值, $|ACE|$ 表示ACE绝对值的平均值. 从图11中可以看出, 同其他算法相比, GWDQ频率偏差减少了6.22%~68.94%, 稳态误差减少了12.76%~50.95%, ACE减少了36.08%~75.2%, 超调量减少了7.75%~37.2%.

4.2 多区域互联的东北电网模型

到2014年末, 辽宁电网装机容量达到22370 MW^[24],

图 9 阶跃扰动下不同算法的控制性能. (a) 阶跃扰动结果图; (b) 阶跃扰动性能指标

Figure 9 Control performance of different algorithms under step disturbance. (a) The result graph of step disturbance; (b) the performance index of step disturbance.

图 10 阶跃扰动下不同算法的收敛情况
Figure 10 Convergence of different algorithms under step disturbance.

图 11 方波扰动下不同算法的控制性能
Figure 11 Control performance of different algorithms under square wave disturbance.

其中火电占比80%, 风电占比20%; 黑龙江电网装机容量达到2503.3 WM^[25], 其中火电占比77.53%, 水电占比3.91%, 风电占比18.52%; 吉林电网装机容量为15588 MW^[26], 其中火电占比84.91%, 风电占比13.28%, 水电占比1.03%, 生物质能占比0.80%. 根据以上三省数据, 本文搭建的东北电网模型以南方电网等值模型为基础, 融入了包括水电、火电、柴油发电、生物质能、冷热电联产在内的多种能源类型, 共计78台机组, 实现多区域能源互联.

对基于东北电网的AGC控制系统, 施加幅值1500 MW的随机阶跃扰动. 以10000 s的随机阶跃扰动作为考核周期, 测试Q, DQL, PDWOLF-PHC(λ), DDRQN, DDQN-AD, GWO, PRDDQN-AD, GWDQ 8种算法的性能. 图12为控制器输出, 表2为8种算法的控制性能, 相比较于其他算法, 所提算法的 $|\Delta f|$ 降低了34.55%~48.67%, ACE降低了19.93%~56.69%, CPS1提高了2.64%~41.03%, 稳态误差降低了12.32%~49.66%.

5 结论

针对传统强化学习无法有效解决大规模新能源接入所带来的强随机扰动, 以至于AGC性能变差的问题, 提出一种新颖且高效的GWDQ策略, 来快速获取AGC过程中多区域协同最优解.

所提策略是以DQL算法为支点且融入了GWO算法, 能够解决基于传统Q框架体系的这一类强化学习算法中, 状态动作对值高估问题的同时, 在未知的搜

图 12 随机阶跃扰动下不同算法的控制性能
Figure 12 Control performance of different algorithms under random step disturbance.

表 2 不同算法下的控制性能指标

Table 2 Control performance under different algorithms

算法	$ \Delta f $ (Hz)	ACE (MW)	CPS1 (%)	稳态误差 (%)
GWDQ	0.003773	0.8142	199.9974	0.8180
PRDDQN-AD	0.005765	1.0168	194.7118	0.9329
GWO	0.006282	1.1984	179.9824	0.9993
DDQN-AD	0.006253	1.3395	176.7471	0.9996
DDRQN	0.006382	1.4086	173.9890	1
PDWOLF-PHC(λ)	0.006416	1.6979	165.4626	1.0002
DQL	0.006557	1.7006	142.9738	1.0004
Q	0.007351	1.8801	117.9365	1.6248

索空间中快速搜索最优解。

通过对改进的IEEE两区综合能源系统模型以及多区域互联的东北电网模型进行仿真, 结果显示, 由于GWDQ能够解决状态动作值高估问题, 从而解决智能体易陷入局部优化问题, 与PRDDQN-AD, DDQN-AD, DDRQN, PDWOLF-PHC(λ), DQL, Q这一类基于传统Q框架体系的强化学习算法相比, 所提策略表现出了较强的在线学习能力和决策能力, 无论在 $|\Delta f|$, ACE,

CPS1还是稳态误差, 性能均是相对更优。而与不具有自学习特性的GWO相比, 所提策略尤其在 $|\Delta f|$, ACE性能优势更加明显。不同工况下的大规模仿真显示, 所提策略能够快速获取分布式AGC模式下的多区域协同最优解。然而区域内机组的功率分配采用实际工程中常用的可调容量等比例分配法, 后续的研究工作将引入智能方法(如一致性Q学习算法), 对机组功率进行动态优化分配, 实现AGC系统整体的智能化。

参考文献

- 1 Yu X D, Xu X D, Chen S Y, et al. A brief review to integrated energy system and energy internet (in Chinese). *J Electr Tech*, 2016, 31: 1–13 [余晓丹, 徐宪东, 陈硕翼, 等. 综合能源系统与能源互联网简述. *电工技术学报*, 2016, 31: 1–13]
- 2 Liu X R. Analysis and prospect of multi-energy complementary system (in Chinese). *Energy Saving*, 2018, 37: 28–33 [刘秀如. 多能互补集成优化系统分析与展望. *节能*, 2018, 37: 28–33]
- 3 Cui K L, Zhang Y F. Classical control and intelligent control for the time delay system (in Chinese). *Microcomput Inform*, 2004, 20: 25–26 [崔坤林, 张翼飞. 时滞系统的经典控制与智能控制. *微计算机信息*, 2004, 20: 25–26]
- 4 Zribi M, Al-Rashed M, Alrifai M. Adaptive decentralized load frequency control of multi-area power systems. *Int J Electr Power Energy Syst*, 2005, 27: 575–583
- 5 Shayeghi H, Shayanfar H A. Decentralized robust AGC based on structured singular values. *Electr Eng*, 2006, 57: 305–317
- 6 Yu T, Zhou B, Chen J R. Stochastic optimal CPS control for interconnected power grids using multi-step backtrack Q(λ) learning (in Chinese). *J Electr Tech*, 2011, 26: 179–186 [余涛, 周斌, 陈家荣. 基于多步回溯Q(λ)学习的互联网随机最优CPS控制. *电工技术学报*, 2011, 26: 179–186]
- 7 Yu T, Yuan Y. An average reward model based whole process R(λ)-learning for optimal CPS control (in Chinese). *Power Syst Autom*, 2010, 34: 27–33 [余涛, 袁野. 基于平均报酬模型全过程R(λ)学习的互联网CPS最优控制. *电力系统自动化*, 2010, 34: 27–33]
- 8 Xi L, Yu L, Fu Y M, et al. Automatic generation control based on deep reinforcement learning with exploration awareness (in Chinese). *China J Electr Eng*, 2019, 39: 4150–4162 [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: 4150–4162]
- 9 Xi L, Chen J F, Huang Y H, et al. Smart generation control based on deep reinforcement learning with the ability of action self-optimization (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]
- 10 Yu T, Zhou B. Reinforcement learning based CPS self-tuning control methodology for interconnected power systems (in Chinese). *Power Syst Prot Control*, 2009, 37: 33–38 [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: 33–38]
- 11 van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double Q-learning. 2015. arXiv: [1509.06461](https://arxiv.org/abs/1509.06461)
- 12 Mirjalili S, Mirjalili S M, Lewis A. Grey wolf optimizer. *Adv Eng Softw*, 2014, 69: 46–61
- 13 Wang B. The world's first battery and gas turbine hybrid energy storage system launched (in Chinese). *Energy Res Inform*, 2016, 32: 231 [王波. 世界上第一款电池和燃气轮机混合储能系统推出. *能源研究与信息*, 2016, 32: 231]
- 14 Pan Z N, Zhang X S, Yu T, et al. Hierarchical real-time optimized dispatching for large-scale clusters of electric vehicles (in Chinese). *Power Syst Autom*, 2017, 41: 96–104 [潘振宁, 张孝顺, 余涛, 等. 大规模电动汽车集群分层实时优化调度. *电力系统自动化*, 2017, 41: 96–104]
- 15 Yu T, Zhou B, Chen J R. Q-learning based dynamic optimal CPS control methodology for interconnected power systems (in Chinese). *Chinese J Electr Eng*, 2009, 29: 13–19 [余涛, 周斌, 陈家荣. 基于Q学习的互联网动态最优CPS控制. *中国电机工程学报*, 2009, 29: 13–19]
- 16 Meng Y, Tu S S, Yu J L. Detection scheme of impersonation attack based on DQL algorithm in fog computing (in Chinese). *Comput Eng Appl*, 2020, 56: 63–68 [孟远, 涂山山, 于金亮. 雾计算中基于DQL算法的伪装攻击检测方案. *计算机工程与应用*, 2020, 56: 63–68]
- 17 Liu R, Lin W, Li W D, et al. Discussion on the evaluation criteria of interconnected power system operation performance (in Chinese). *Autom*

- Electr Power Syst, 2005, 29: 87–91 [刘娆, 林伟, 李卫东, 等. 互联电力系统运行控制性能评价标准的探讨. 电力系统自动化, 2005, 29: 87–91]
- 18 Li Z, Jing D, Zhao Q, et al. Application of AGC control strategy based on CPS/DCS standards in interconnected power system (in Chinese). J Power Syst Autom, 2003, 15: 27–32 [李正, 敬东, 赵强, 等. CPS/DCS标准在大区互联电网AGC控制策略中的应用. 电力系统及其自动化学报, 2003, 15: 27–32]
- 19 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 [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]
- 20 Xi L, Li Y D, Huang Y H, et al. Smart generation control based on the virtual wolf pack control strategy (in Chinese). Chinese J Electr Eng, 2018, 38: 2966–2979 [席磊, 李玉丹, 黄悦华, 等. 基于虚拟狼群控制策略的智能发电控制. 中国电机工程学报, 2018, 38: 2966–2979]
- 21 Xi L, Zhang L, Huang Y H, et al. Multiple level automatic generation control based on the greedy strategy (in Chinese). J China Electr Eng, 2020, 40: 5204–5217 [席磊, 张乐, 黄悦华, 等. 基于贪婪策略的多层自动发电控制. 中国电机工程学报, 2020, 40: 5204–5217]
- 22 Lu X J, Yi J W, Li Y. Optimal control strategy of AGC with participation of energy storage system based on multi-objective mesh adaptive direct search algorithm (in Chinese). Grid Tech, 2019, 43: 2116–2124 [路小俊, 伊建伟, 李炎. 基于多目标网格自适应搜索算法的储能系统参与AGC优化控制策略. 电网技术, 2019, 43: 2116–2124]
- 23 Liao X M, Tong J P, Yu T. Use of microturbine generation system for power system black-start (in Chinese). Power Syst Prot Control, 2009, 37: 83–88 [廖旭明, 童家鹏, 余涛. 微型燃气轮机发电系统在城市电网黑启动中的运用. 电力系统保护与控制, 2009, 37: 83–88]
- 24 Liu H B. Research on the improvement of auxiliary service mechanism and transaction optimization model of wind power peak regulation in Northeast Power Grid (in Chinese). Dissertation for Doctoral Degree. Beijing: North China Electric Power University, 2016 [刘海波. 东北电网风电调峰辅助服务机制完善及交易优化模型研究. 博士学位论文. 北京: 华北电力大学, 2016]
- 25 Zhao J. Influencing factors of wind power absorptive capacity in Heilongjiang grid and improving research measures (in Chinese). Northeast Power Tech, 2017, 38: 21–23 [赵君. 黑龙江电网风电消纳能力影响因素及提高措施研究. 东北电力技术, 2017, 38: 21–23]
- 26 Wang S R, Fu J Y, Liu L, et al. Analysis of primary frequency regulation in Jilin power grid (in Chinese). Jilin Power, 2010, 38: 1–4 [王绍然, 傅吉悦, 刘乐, 等. 吉林电网一次调频分析. 吉林电力, 2010, 38: 1–4]

The multiagent cooperative AGC strategy for multiregional energy interconnection

XI Lei, LIU ZongJing, SUN MengMeng & WU JunNan

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

Traditional reinforcement learning cannot effectively solve the strong random disturbance problem caused by large-scale new energy access; therefore, the response speed of AGC (automatic generation control) becomes slower and the performance becomes poorer. This study considers the double Q learning that can solve the problem of overestimation of the value of state-action pairs in the reinforcement learning algorithm based on the Q framework as the fulcrum. It integrates a grey wolf optimization algorithm that can quickly search for the optimal solution in the unknown search space. Additionally, it proposes a multiagent cooperative AGC strategy for multiregional energy interconnection, namely, the GWDQ strategy, to quickly obtain the multiregional cooperative optimal solution in the AGC process. The two-area integrated energy system model, such as the hybrid-electric gas turbine system, CCHP and other forms of energy, and the northeast power grid model with a multiregional energy interconnection, are simulated. The results showed that the proposed strategy has a strong learning ability. Besides, the convergence speed and control performance are superior to that of the traditional reinforcement learning algorithm. The proposed strategy can quickly obtain the multiregional collaborative optimal solution in the AGC process.

double Q learning, grey wolf optimization, AGC, multi-agent

doi: [10.1360/SST-2020-0467](https://doi.org/10.1360/SST-2020-0467)