

# 分布式多区域多能微网群协同 AGC 算法

席磊<sup>1</sup> 周礼鹏<sup>1</sup>

**摘要** 综合能源多区域协同是电网发展趋势,而核心问题是采用何种方法对多区域进行协同.本文基于  $Q(\sigma)$  融入了资格迹及双重  $Q$  学习,提出一种面向多区域多能微网群的多智能体协同控制算法,即  $DQ(\sigma, \lambda)$ ,避免传统强化学习动作探索值高估的同时,来获取分布式多区域的协同.通过对改进的 IEEE 两区域负荷频率控制模型及三区域多能微网群自动发电控制(Automatic generation control, AGC)模型仿真,结果表明,与传统方法相比,所提算法具有快速收敛性和更优动态性能,能获得分布式多区域多能微网群的协同.

**关键词** 综合能源,多能微网,自动发电控制,强化学习,双重  $Q$  学习

**引用格式** 席磊,周礼鹏.分布式多区域多能微网群协同 AGC 算法.自动化学报,2020,46(9):1818-1830

**DOI** 10.16383/j.aas.c200105

## Coordinated AGC Algorithm for Distributed Multi-region Multi-energy Micro-network Group

XI Lei<sup>1</sup> ZHOU Li-Peng<sup>1</sup>

**Abstract** Comprehensive energy multi-region coordination is the development trend of the power grid, and the core question is what method to use for multi-region coordination. Based on the integration of the qualification trace and dual  $Q$ -learning in  $Q(\sigma)$ , this paper proposes a multi-agent collaborative control algorithm for multi-region and multi-energy micro-grid group, named  $DQ(\sigma, \lambda)$ , to avoid high exploration value of traditional reinforcement learning actions. At the same time of evaluation, the distributed multi-region collaboration is obtained. Simulations of the improved IEEE two area load frequency control model and the three area multi-energy microgrid group automatic generation control (AGC) model show that the proposed algorithm has fast convergence and better dynamic performance than traditional methods, and can achieve distributed Synergy of regional multi-energy microgrid groups.

**Key words** Comprehensive energy, multi-energy microgrid, automatic generation control (AGC), reinforcement learning, double  $Q$ -learning

**Citation** Xi Lei, Zhou Li-Peng. Coordinated AGC algorithm for distributed multi-region multi-energy micro-network group. *Acta Automatica Sinica*, 2020, 46(9): 1818-1830

发展新能源能够解决化石燃料燃烧引起的环境恶化问题,集成了源、荷、气、热、储等多种分布式能源<sup>[1]</sup>的综合能源系统<sup>[2-3]</sup>势在必行,但规模化的分布式新能源并网将带来强随机扰动,以及由于传统机组惯性降低、缺乏辅助频率支持、调频容量不足等引起的频率失稳问题<sup>[4]</sup>,给现代电力系统的运行和控制提出了新的挑战.因此,本文从自动发电控制(Automatic generation control, AGC)角度面向多区域多能微网群提出一种新的频率控制方法以实现

多区域协同控制.

当前 AGC 控制方法主要分为传统解析式和机器学习两大类.基于传统解析式的控制方法,以 PID 控制方法为代表<sup>[5-6]</sup>.文献 [7] 提出了基于灰狼优化算法的分数阶 PID 控制器参数优化整定方案,解决了网络化时滞互联电网的负荷频率控制 (Load frequency control, LFC) 问题.文献 [8] 提出了一种基于社会学习自适应细菌觅食算法的最优 PI/PID 控制器设计方法,以解决互联电网 AGC 控制器参数优化整定问题.文献 [9] 提出了基于随机帝国竞争算法的级联模糊分数阶 CFFOPI-FOPID 控制器,以解决 AGC 问题.传统控制方法主要根据区域控制偏差误差 (Area control error, ACE) 单一化地确定总调节功率,控制机组出力.然而电力系统新形态下区域间互动变化灵活,需要根据长期历史数据进行学习、分析、存储,以对多区域进行协同控制<sup>[10]</sup>.

收稿日期 2020-03-05 录用日期 2020-04-27

Manuscript received March 5, 2020; accepted April 27, 2020

国家自然科学基金 (51707102) 资助

Supported by National Natural Science Foundation of China (51707102)

本文责任编辑 孙秋野

Recommended by Associate Editor SUN Qiu-Ye

1. 三峡大学电气与新能源学院 宜昌 443002

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

随着人工智能的崛起, 一些学者将人工智能方法应用于 AGC, 试图解决上述问题. 基于人工智能的强化学习能够通过与环境探索试错积累经验分析获取最优策略, 机器学习体系应用在 AGC, 尤以基于强化学习的 Q 学习应用最为广泛. 文献 [11] 基于 Q 学习提出了一种改进的极限 Q 学习算法, 对微电网的下垂控制进行参数整定, 从而实现频率调节与经济调度的一体化. 文献 [12] 提出了一种孤岛运行模式下基于平均报酬模型的多步  $R(\lambda)$  算法的 AGC 控制器, 以实现微电网的智能发电控制与频率调整. 文献 [13] 提出了一种基于模型预测控制的孤岛微电网频率二次控制策略, 以解决不确定延时对系统频率的影响, 并采用小信号模型和参与因子分析系统的稳定性. 然而上述文献为单区域模型, 同样算法也为单智能体算法, 这种无多区域协同的模式, 可等效看作“集中式”控制, 无法满足日益发展的综合能源模式下分布式多区域协同发展趋势.

多智能体强化学习是解决多智能体系统问题的一种有效方法, 而协作多智能体强化学习专注于解决协作问题. 协作多智能体强化学习与分布式优化有非常密切的联系, 因此求解分布式优化的高效最优方法可以引入求解协作多智能体强化学习问题<sup>[4]</sup>. 文献 [15] 针对多区域互联微网系统, 结合线性自抗扰控制算法和基于原对偶梯度算法的多智能体系统, 提出了一种新的分布式优化控制算法, 有效地结合系统动态特性与优化过程解决负荷频率控制问题. 文献 [16] 在微网分层控制结构的框架下, 提出多智能体自适应控制算法, 使频率恢复额定值, 且有功功率按各分布式电源的额定功率比例分配. 文献 [17] 在 Q 学习基础上提出了一种面向混合交互环境的基于多智能体系统 (Multi-agent system, MAS) 和元胞自动机的微网分布式协调自趋优控制策略, 调节微源的有功和无功出力及系统频率. 文献 [18] 提出一种基于多智能体微电网控制框架的多智能体协作学习算法, 有效管理微网中的微电源促使微网协调控制. 文献 [19] 面向分布式能源提出一种基于虚拟狼群控制策略的分层分布式控制—PDWoLF-PHC ( $\lambda$ ), 算法中融入资格迹<sup>[20]</sup>, 能够解决算法的时间信度分配问题, 以提高算法收敛速度, 进而来获得区域的最优控制. 然而上述的控制算法均为基于传统强化学习算法, 此类算法在随机环境中容易出现动作值在探索过程中的“高估”现象, 会导致决策质量低. 且上述算法均属于离策略, 其面临的主要问题是离策略算法难以收敛、收敛速度慢以及收敛精度低.

因此, 为解决上述问题, 通过引入参数  $\sigma$  统一

离策略与在策略的优缺点, 提出了基于“将各种看似不同的算法思想联合统一以产生更好的算法”思想的  $Q(\sigma)$  算法<sup>[21]</sup>. 为解决随机环境中传统强化学习算法的高估动作值, 提高算法收敛速度, 以实现多区域电力系统协同控制, 本文根据协作多智能体强化学习在  $Q(\sigma)$  算法基础上融入资格迹与双重学习<sup>[2]</sup>, 提出了一种基于多步统一强化学习的多智能体协同  $DQ(\sigma, \lambda)$  控制算法. 算法中固有的偏差与方差权衡主要取决于参数  $\sigma$ , 当  $\sigma = 0$  时,  $DQ(\sigma, \lambda)$  处于全采样 Double Q ( $\lambda$ ) 算法; 当  $\sigma = 1$  时,  $DQ(\sigma, \lambda)$  处于纯期望 Double Expected-Sarsa ( $\lambda$ ) 算法; 当  $\sigma = 0.5$  时,  $DQ(\sigma, \lambda)$  处于采样和期望的混合算法. 通过对改进的 IEEE 标准两区域负荷频率控制模型以及分布式三区域多能微网 AGC 模型进行仿真, 验证所提算法的有效性.

## 1 $DQ(\sigma, \lambda)$ 算法

由于传统强化学习过度追求长期折扣回报奖励最大, 在策略往往选择对应最大 Q 值的动作, 使策略探索过程出现动作值的高估, 进而产生累积高偏差, 影响智能体学习到最优策略. 为此, 本文根据协作多智能体强化学习在  $Q(\sigma)$  算法基础上, 融入了资格迹及双重学习, 进而提出一种新型  $DQ(\sigma, \lambda)$ , 通过解决传统强化探索过程中动作值高估问题, 进而获得分布式多区域多能微网群的协同控制.

### 1.1 $Q(\sigma)$ 学习算法

时间差分学习 (Temporal-difference learning, TD)<sup>[23]</sup> 是无模型强化学习中最重要策略之一, TD 方法结合了蒙特卡罗方法和动态规划的优点, 适用于无模型、持续进行的任务. 常见 TD 有 Q、Sarsa、Expected-Sarsa<sup>[24]</sup> 等, 其中 Q 学习应用最广泛. 同样, 文献 [21] 通过引入采样参数  $\sigma$ , 统一了 Sarsa 算法 (全采样) 和 Expected-Sarsa 算法 (纯期望), 提出了一种统一在策略与离策略的 TD 算法, 即  $Q(\sigma)$ . 其中, 离策略和在策略主要区别是在策略一般只有一个策略 (常用  $\epsilon$  贪婪策略). 而离策略一般有两个策略, 行为策略 (常用  $\epsilon$  贪婪策略) 用于选择新的动作, 目标策略 (常用 Max 贪婪法) 用于更新价值函数. 本文所有提及算法所涉及策略均为上述常用策略.

Sarsa 是一种经典的在策略 TD 算法, 它将动作值函数作为其估计值, 而非状态值函数. 特别地, 对于在策略算法, 其必须根据当前行为策略与所有状态动作估算最优 Q 值. Sarsa 算法至始至终只使用  $\epsilon$  贪婪策略更新价值函数和选择新的动作, 其更新方式如下:

$$Q_{k+1}(s, a) = Q_k(s, a) + \alpha \delta_k^s \quad (1)$$

$$\delta_k^s = R_{k+1} + \gamma Q_k(s_{k+1}, a_{k+1}) - Q_k(s_k, a_k) \quad (2)$$

其中,  $R_{k+1} + \gamma Q_k(s_{k+1}, a_{k+1})$  称为 TD 目标, 奖励加上下一个状态和下一个动作的折扣值组成。

Expected-Sarsa 作为一种离策略学习算法, 可将 Q 学习算法推广到任意目标策略, 根据目标策略利用下一个状态-动作值对的期望值进行 Q 值估算:

$$Q_{k+1}(s, a) = Q_k(s, a) + \alpha \delta_k^{es} \quad (3)$$

$$\delta_k^{es} = R_{k+1} + \gamma \sum_{a \in A} \pi(s_{k+1}, a) Q_k(s_{k+1}, a) - Q_k(s_k, a_k) \quad (4)$$

其中,  $\delta_k^{es}$  是第  $k$  个预期 TD 误差. 虽然 Expected-Sarsa 在计算上比 Sarsa 更复杂, 但作为回报, 它消除了由于下一个动作随机选择而产生的方差. 当经历相同的探索经验, Expected-Sarsa 的表现优于 Sarsa. 此外, 在步长参数  $\alpha$  的取值范围内, Expected-Sarsa 表现比 Sarsa 有显著改善.

当  $0 < \sigma < 1$  时, 性能优于  $\sigma = 0$  或 1 极端情况, 故 Q ( $\sigma$ ) 算法是通过采样参数  $\sigma$  在 Sarsa ( $\sigma = 1$ , 全采样) 更新和 Expected-Sarsa ( $\sigma = 0$ , 纯期望) 更新之间进行线性加权:

$$Q_{k+1}(s, a) = Q_k(s, a) + \alpha \delta_k^\sigma \quad (5)$$

$$\delta_k^\sigma = R_{k+1} + \gamma [\sigma Q_k(s_{k+1}, a_{k+1}) + (1 - \sigma) \sum_{a \in A} \pi(s_{k+1}, a) Q_k(s_{k+1}, a)] - Q_k(s_k, a_k) \quad (6)$$

式中,  $\delta_k^\sigma$  是经参数  $\sigma$  加权后的 TD 误差.

## 1.2 DQ( $\sigma, \lambda$ ) 算法

以离散时间马尔科夫决策过程为数学基础, 基于 Q ( $\sigma$ ) 算法并融入资格迹, 提出了一种新颖的快速多步算法 Q ( $\sigma, \lambda$ ), 以解决 Q ( $\sigma$ ) 的时间信度分配问题, 进而可提高 AGC 机组功率调节快速性. 其 TD 目标是 Sarsa 和 Expected-Sarsa 的加权, 其中参数  $\sigma$  为控制权重. 当  $\sigma = 0$  时, Q ( $\sigma, \lambda$ ) 的目标等于 Q ( $\lambda$ ) 目标, 因此资格迹更新减少到标准累积资格迹更新. 当  $\sigma = 1$  时, Q ( $\sigma, \lambda$ ) 的目标等于 Expected-Sarsa ( $\lambda$ ) 目标, 资格迹是目标策略当前动作概率的线性加权. 资格迹更新方式为

$$e_k(s, a) = \gamma \lambda e_{k-1}(s, a) \times [\sigma + (1 - \sigma) \pi(a_{k+1} | s_{k+1})] + 1, \\ Q_{k-1}(s_k, a_k) = \max_a Q_{k-1}(s_k, a) \quad (7)$$

$$e_k(s, a) = \gamma \lambda e_{k-1}(s, a) \times [\sigma + (1 - \sigma) \pi(a_{k+1} | s_{k+1})], \quad \text{其他} \quad (8)$$

Q ( $\sigma, \lambda$ ) 的迭代更新式为

$$Q_k(s_k, a_k) = Q_k(s_k, a_k) + \alpha \delta_k^\sigma e_k(s, a) \quad (9)$$

同时, 为了解决本文的核心问题, 即策略探索过程中动作值高估问题, 在 Q ( $\sigma, \lambda$ ) 的基础上采用去耦“动作选择”和“动作评估”相结合的双重学习, 进而形成 Double Q ( $\sigma, \lambda$ ), 下面简称 DQ ( $\sigma, \lambda$ ). 本文中两个不同的值函数  $Q_A$  和  $Q_B$  替代单一值函数  $Q$ , 对  $Q_A + Q_B$  使用行为策略并对动作进行采样, 在每次迭代中随机更新值函数  $Q_A$  与  $Q_B$ :

$$Q_A(s_k, a_k) = Q_A(s_k, a_k) + \alpha \delta_k^A e_k(s, a) \quad (10)$$

$$Q_B(s_k, a_k) = Q_B(s_k, a_k) + \alpha \delta_k^B e_k(s, a) \quad (11)$$

当更新  $Q_A$  或  $Q_B$  时, DQ ( $\sigma, \lambda$ ) 的 TD 误差则按以下方式进行更新:

$$\delta_k^A = R_{k+1} + \gamma [\sigma Q_B(s_{k+1}, a_{k+1}) + \sum_a \pi(a | s_{k+1}) Q_B(s_{k+1}, a)] - Q_A(s_k, a_k) \quad (12)$$

$$\delta_k^B = R_{k+1} + \gamma [\sigma Q_A(s_{k+1}, a_{k+1}) + \sum_a \pi(a | s_{k+1}) Q_A(s_{k+1}, a)] - Q_B(s_k, a_k) \quad (13)$$

其中,  $\delta_k^A$  为更新时  $Q_A$  产生的 TD 误差,  $\delta_k^B$  为更新  $Q_B$  时产生的 TD 误差,  $\gamma$  为折扣因子  $\pi(a | s_{k+1})$  是将状态映射到动作概率的动作函数.

## 2 基于 DQ( $\sigma, \lambda$ ) 算法的 AGC 设计

基于 DQ ( $\sigma, \lambda$ ) 的多能微网群分布式多区域多智能体协同控制架构如图 1 所示, 智能体全面感知源-网-荷-储设备运行信息, 对于多区域互联电网联络线和频率偏差模式下的 AGC, 国内外常用的评估方法是北美电力可靠性委员会提出的 CPS 标准<sup>[2]</sup>.

### 2.1 AGC 奖励函数

本文将以 CPS、ACE、频率偏差作为综合目标函数以获取最优 AGC 机组出力, 进而在最优 AGC 控制策略下达到系统功率平衡. 基于 DQ ( $\sigma, \lambda$ ) 的 AGC 控制器某  $i$  区域的综合奖励函数表示为

$$R_i(k) = \tau_i, \quad \tau_i \geq 0, \quad C_{CPS1i}(k) \geq 200 \quad (14)$$

$$R_i(k) = -\{\omega_{1i} [E_{ACEi}(k) - E_{ACEi}^*]^2 + \mu_{1i} [a_{ord-i}(k) - a_{ord-i}^*]^2\}, \\ C_{CPS1i}(k) \in [100, 200) \quad (15)$$

$$R_i(k) = -\{\omega_{2i} [C_{CPS1i}(k) - C_{CPS1i}^*]^2 + \mu_{2i} [a_{ord-i}(k) - a_{ord-i}^*]^2\}, \\ C_{CPS1i}(k) < 100 \quad (16)$$

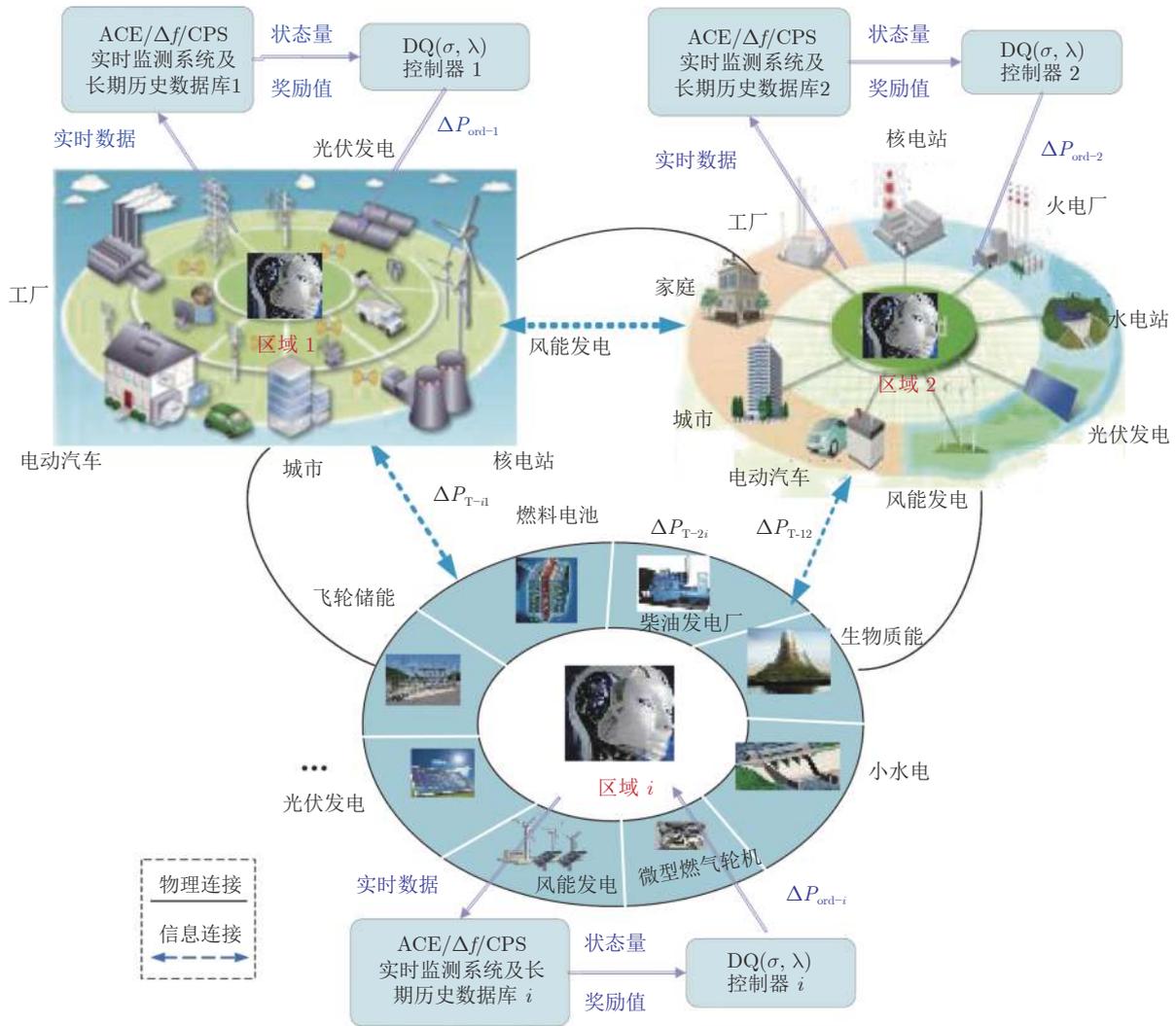


图 1 多能微网群多区域协同控制架构

Fig.1 Multi-energy microgrid group multi-region cooperative control architecture

其中,  $\tau_i$  为任意非负数, 本文取 0;  $C_{CPS1i}(k)$  与  $E_{ACEi}(k)$  分别为 CPS1 和 ACE 在第  $k$  步迭代时刻的瞬时值;  $a_{ord-i}(k)$  为  $k$  时刻的控制动作集  $A$  的指针;  $a_{ord-i}^*$  即功率控制动作作为 0 时的指针, 引入动作变化项, 是为了限制控制器输出功率指令频繁大幅度升降引起的系统振荡和经济代价;  $\omega_{1i}, \omega_{2i}$  和  $\mu_{1i}, \mu_{2i}$  分别为状态输入和控制动作的优化权值, 相当于线性二次型调节器控制性能指标中的 Q 和 R 权值参数<sup>[26]</sup>;  $C_{CPS1i}^*$  为 CPS1 指标控制期望值;  $E_{ACEi}^*$  为 ACE 控制期望值。

### 2.2 参数设置及算法流程

AGC 控制系统的设计需要对系统参数进行合理的设置, 其中:

1) 学习率  $\alpha$  ( $0 < \alpha < 1$ ),  $\alpha$  表示要给改善的算

法更新部分多少信任度, 较大的  $\alpha$  值会加快  $DQ(\sigma, \lambda)$  算法的收敛速度, 而较小的  $\alpha$  值能保证控制器的搜索空间, 从而提高  $DQ(\sigma, \lambda)$  算法收敛的稳定性, 本文  $\alpha$  取为 0.1。

2) 折扣因子  $\gamma$  ( $0 < \gamma < 1$ ), 函数的未来奖励的衰减值, 当  $\gamma$  趋向于 1 时, 考虑长期奖励, 当  $\gamma$  趋近于零时, 只能看到当前奖励. 本文  $\gamma$  取为 0.9。

3) 资格迹衰退系数  $\lambda$  ( $0 < \lambda < 1$ ), 其主要作用是在状态-动作对中分配信誉, 影响收敛速度, 本文  $\lambda$  取为 0.95。

4) 控制采样权重  $\sigma$  ( $0 < \sigma < 1$ ), 具有中等  $\sigma$  的  $Q(\sigma)$  可以胜过  $Double\ Q(\lambda)$ 、 $Double\ Expected\ Sarsa(\lambda)$  和  $Double\ Sarsa(\lambda)$  算法. 本文参数  $\sigma$  取为 0.5。

5) 探索率  $\varepsilon$  ( $0 < \varepsilon < 1$ ), 策略以  $1-\varepsilon$  的概率选择当前最大值的动作, 以  $\varepsilon$  的概率随机选择新动作. 本文仿真预学习时  $\varepsilon$  取 0.5, 在线运行时  $\varepsilon$  取 0.9. DQ ( $\sigma, \lambda$ ) 的算法流程如图 2 所示.

### 3 仿真研究

#### 3.1 改进的 IEEE 标准两区域负荷频率控制模型

将电池储能系统 (Battery energy storage system, BESS)<sup>[27-28]</sup> (如图 3) 融入 IEEE 标准两区域负荷频率控制模型<sup>[29]</sup>, 改进后的模型如图 4 所示. 其中,  $P_{\text{ref}}$  为储能系统目标指令;  $P_{\text{req}}$  为经过并网能量转换系统和响应延时后的功率需求;  $T_{\text{PCS}}$ 、 $T_{\text{DB}}$  分别为并网能量转换环节和响应延时-时间转换环节的时间常数;  $T_s$  为仿真时间与实际时间关系的转换系数;  $\Delta P_G$ ,  $P_{\text{BESS}}$ ,  $S_{\text{SOC}}$  分别为经过电池储能电源

后的有功功率、实际的有功功率、荷电状态;  $K_T$  为积分电量计算时间常数, 其值与仿真时间步长相关;  $E_B$  为储能系统额定容量;  $S_{\text{SOCinit}}$  为储能荷电状态初始值;  $f(S_{\text{SOC}})$  描述了储能输出功率与荷电状态的映射关系.  $B_i$  为各区域的频率偏差因子,  $\Delta P_G$  为发电机输出功率,  $T_g$  为调速器时间常数,  $T_t$  为汽轮机组时间常数,  $T_p$  为频率响应等效函数系数,  $T_s$  为二次时间延时,  $K_p$  为频率响应等效函数系数,  $T_{12}$  为联络线时间常数,  $\Delta P_{\text{tie}}$  为联络线交换功率. 采样周期为 4 s,  $T_{\text{PCS}} = 0.01$  s,  $T_{\text{DB}} = 0$  s,  $T_s = 1$  s; 荷电状态限幅环节的控制区间为 (10, 90); 荷电状态理想运行区间设定为 [30, 70], 仿真初始储能荷电状态为 50%.

在线运行之前, DQ ( $\sigma, \lambda$ ) 需进行充分的预学习, 即通过随机探索试错训练, 使控制器收敛于最优策略  $\pi^*$ , 再投入到真实仿真环境参与在线优化运行. 而对于多能微网群的控制性能, 可通过 CPS、联络线交换功率偏差  $P_{\text{tie}}$ 、频率偏差  $\Delta f$  (合格范围  $\pm(0.05 \sim 0.2)$  Hz) 进行评估. 为兼顾互联区域的频率稳定性和本区域的经济性, 取 CPS1 接近而不是大于 200% 为最优. CPS 标准具体如下:

- 1) 若  $\text{CPS1} \geq 200\%$ , 且  $\text{CPS2}$  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.

##### 3.1.1 正弦负荷离线预学习

在预学习阶段, 引入正弦负荷扰动 (周期 1 200 s, 幅值 1 000 kW, 时长 10 000 s), 对 DQ ( $\sigma, \lambda$ ) 进行训练探索使其收敛于最优策略. 图 5 给出了 DQ ( $\sigma, \lambda$ ) 控制器负荷扰动下两区域预学习性能指标, 由图 5(a) 可见两区域 DQ ( $\sigma, \lambda$ ) 控制器在 2 000 s 以内基本可以跟踪负荷扰动曲线. 图 5(b) 为联络线交换功率偏差变化曲线, 两区域  $P_{\text{tie}}$  (交换功率偏差绝对值的平均值) 为 1.2255 kW. 图 5(c) 是扰动下的频率变化曲线, 两区域的  $\Delta f_{\text{max}}$  (最大频率偏差) 分别为 0.049 Hz 和 0.055 Hz, 远小于实际工程要求 0.2 Hz, 可见控制器具有较强的稳定性. 图 5(d) 为两区域  $E_{\text{AVE-10-min}}$  (10 min ACE 的平均值) 的学习曲线, 其值分别为 1.7214 kW、1.9864 kW, ACE 的 10 min 考核指标值始终保持在 2 kW (2%) 以内. 图 5(e) 为 A、B 两区域  $\text{CPS1}_{\text{AVE-10-min}}$  (10 min CPS1 的平均值) 在学习过程的变化曲线, A、B 区域的  $\text{CPS1}_{\text{AVE-10-min}}$  分别是 199.4934%、199.2681%, CPS1 的 10 min 考核指标值保持在 185% 以上.

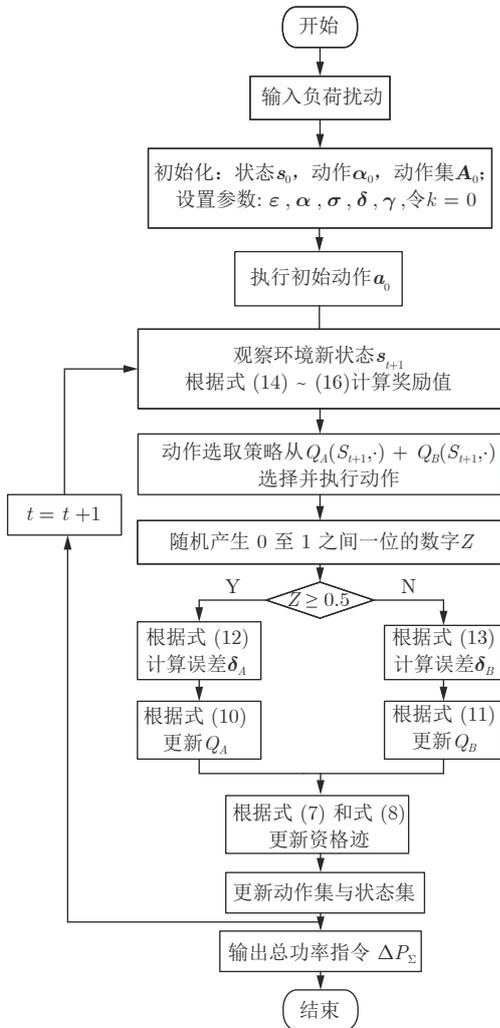


图 2 DQ ( $\sigma, \lambda$ ) 的算法流程

Fig. 2 Algorithm flow of DQ ( $\sigma, \lambda$ )

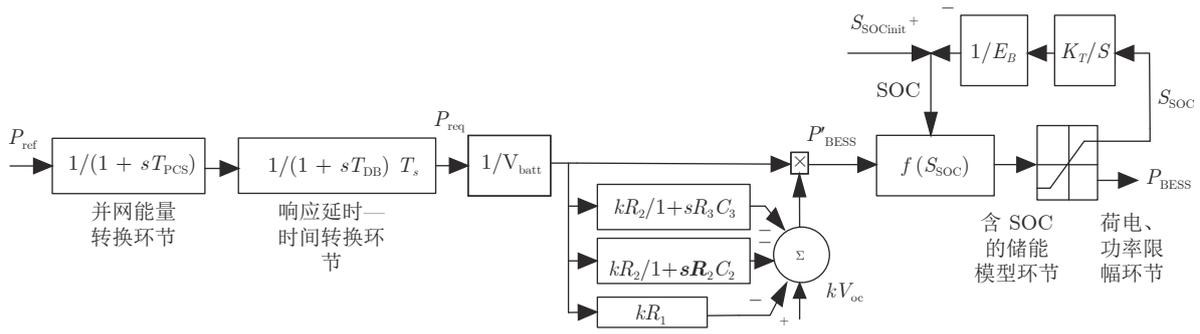


图 3 BESS 仿真模型

Fig.3 BESS simulation model

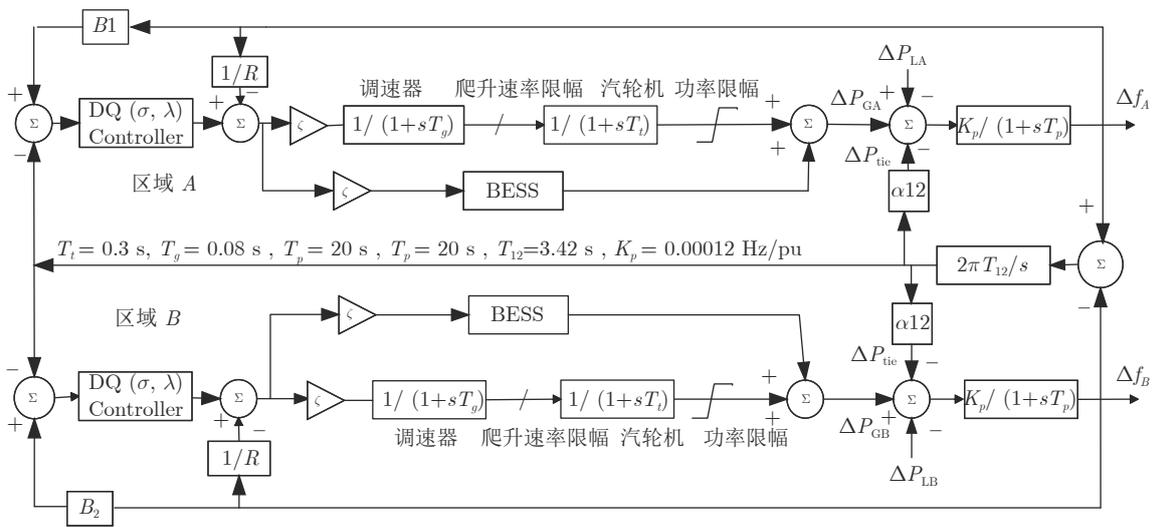


图 4 改进的 IEEE 标准两区域负荷频率控制模型

Fig.4 Improved IEEE standard two-area load frequency control model

另外, 在最优策略中, 选择 2 范数  $\|Q_k(s, a) - Q_{k-1}(s, a)\|^2 \leq \epsilon$  ( $\epsilon = 0.0001$  为指定标准) 作为预学习达到最优策略的终止标准<sup>[30]</sup>, 图 5(f) 中为预学习期间 A 区域 Q 函数差分的收敛结果.  $DQ(\sigma, \lambda)$  收敛于第 255 步 (仿真步长为 4, 即预学习需 1004 s). 基于全球微波互联无线通信技术, 每次信息传输和迭代计算所需时间为 1 ms, 即系统的调节时间为 1.004 s, 满足实际电网 AGC 系统 4 s 的时间尺度要求. 图 5(g) 为引入  $Q, Q(\lambda), Q(\sigma), PDWoLF-PHC(\lambda)$  智能算法的对比收敛效果, 由图可知  $DQ(\sigma, \lambda)$  算法可提高收敛速度 93.92% ~ 98.98%. 综上所述, 在经过大量的训练探索后,  $DQ(\sigma, \lambda)$  控制器已逼近确定性最优 CPS 控制策略, 可将  $DQ(\sigma, \lambda)$  控制器投入真实环境运行.

### 3.1.2 阶跃、随机白噪声负荷在线运行

在线运行时, 对两区域模型引入阶跃负荷扰动, 模拟大规模随机扰动情况. 对  $Q, Q(\lambda), Q(\sigma), PDWoLF-PHC(\lambda), DQ(\sigma, \lambda)$  五种算法的控制器引入了时长 9 000 s, 幅值 1 000 kW 的阶跃负荷扰动进行仿真对比分析. 图 6 为 A 区域分别基于 5 种智能算法的 AGC 控制器的控制性能指标, 图 6(a) 是联络线交换功率偏差变化曲线, 各算法  $P_{tie}$  分别为 9.7430 kW、1.5367 kW、0.6725 kW、0.6296 kW、0.4514 kW,  $DQ(\sigma, \lambda)$  控制器所产生的交换功率偏差最小. 图 6(b) 是 5 种算法频率变化曲线对比效果图, 各算法  $|\Delta f|$  分别为 0.0047 Hz、0.0016 Hz、0.0014 Hz、0.0014 Hz、0.0008 Hz, 相较于其他算法,  $DQ(\sigma, \lambda)$  的  $|\Delta f|$  降低了 42.85% ~ 82.97%. 图 6(c)

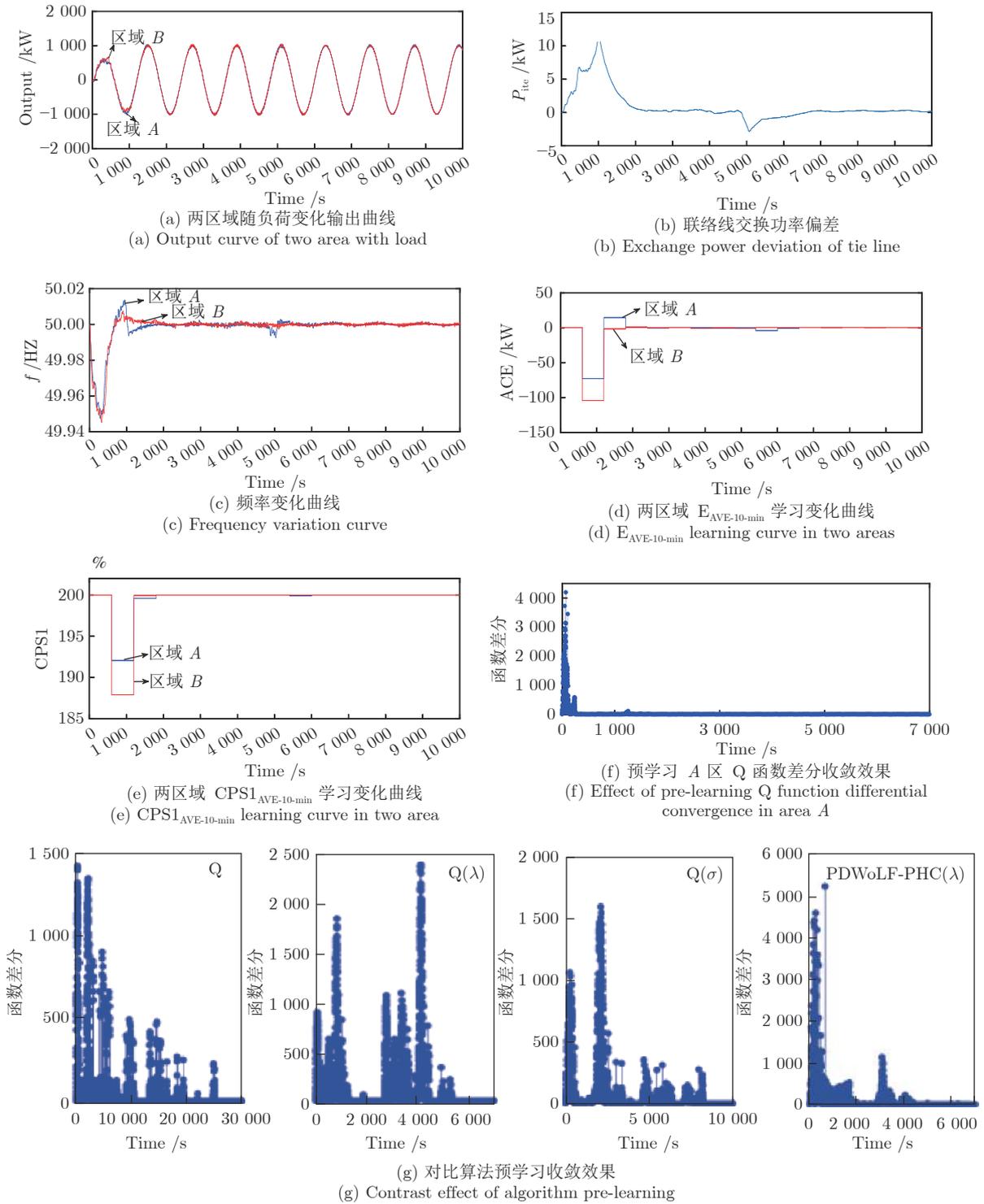


图 5 两区域预学习效果及收敛效果

Fig.5 Pre-learning and convergence effect in two area

为 10 min ACE 的平均值, 各算法值分别为 19.8983 kW、4.4539 kW、3.3934 kW、3.4009 kW、2.5961 kW,  $DQ(\sigma, \lambda)$  的 ACE 降低了 23.66%~86.95%。图 6(d) 为 10 min CPS1 的平均值变化曲

线, 5 种算法值分别是 199.4404%、199.8431%、199.8849%、199.8998%、199.9122%,  $DQ(\sigma, \lambda)$  的 CPS1 提高了 0.0062%~0.2365%。

为考虑更加实际的运行情况, 进一步验证所提

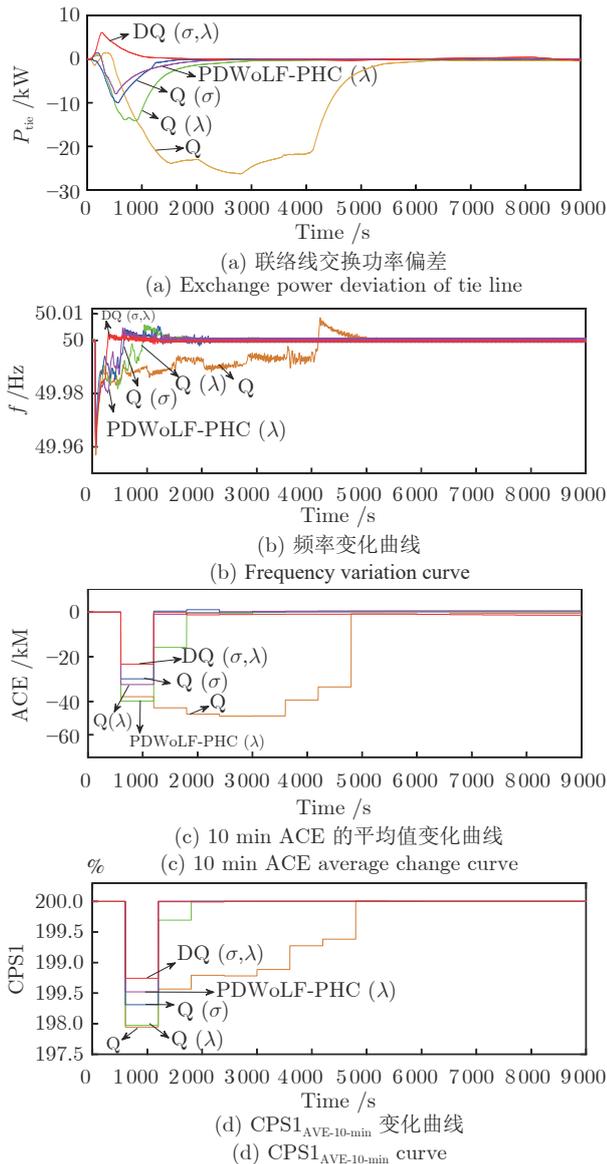


图 6 阶跃负荷扰动下不同算法的性能指标

Fig. 6 Performance index of different algorithms under step load disturbance

算法的控制性能. 通过引入随机白噪声 (噪声功率: 10000 kW; 检测周期: 60 s; seed: 23341) 作为负荷扰动, 模拟接入未知的分布式新能源的随机负荷扰动情况, 以 5 小时的负荷扰动为考核周期, 测试  $DQ(\sigma, \lambda)$ 、 $PDW\omega LF-PhC(\lambda)$ 、 $Q(\sigma)$ 、 $Q(\lambda)$ 、 $Q$  等 5 种控制器的性能. 图 7 为上述 5 种智能算法在 A 区域的各项性能考核指标,  $DQ(\sigma, \lambda)$  能降低  $|\Delta f|$  28.17% ~ 57.73%, 减少  $|ACE|$  6.63% ~ 33.85%. 仿真结果表明,  $DQ(\sigma, \lambda)$  在能源出力不确定、负荷随机波动的情况下仍然能保持稳定的控制效果.

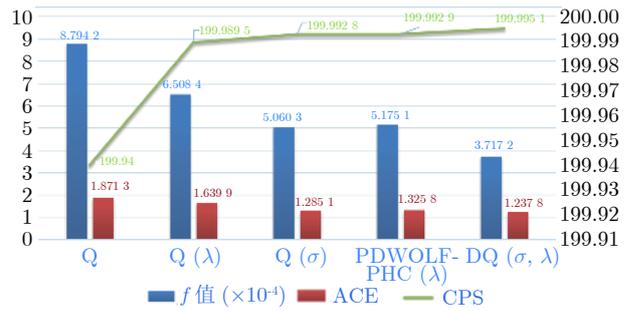


图 7 随机白噪声扰动下不同算法的控制性能

Fig. 7 Control performance of different algorithms under stochastic white noise disturbance

### 3.2 分布式多区域多能微网群协同 AGC 模型

综合能源系统作为能源互联网的物理载体, 能够利用其多能互补的优势, 对不同类型的能源进行协调管理和分配, 在满足用户多种能源需求的同时, 进一步减少温室气体排放、提高能源综合利用率和降低能源供应成本. 在传统模式下, 多区域综合能源系统由于地理位置分散, 往往都是独立运行, 彼此间缺乏协调控制, 容易存在资源配置不合理等问题.

因此, 本文搭建了融入大量新能源的分布式多区域 (以 3 区域为例) 多能微网群协同 AGC 模型, 以验证  $DQ(\sigma, \lambda)$  的实际工程应用效果. 模型中包括光伏、风电、小水电、微型燃气轮机、柴油发电机储、生物质能、燃料电池<sup>[31-33]</sup>, 其拓扑结构如图 8, 模型参数如表 1. 其中 Area 1 和 Area 3 模型参数和机组参数相同, 3 区域的调节功率分别为 2 350 kW、2 590 kW 和 1 840 kW, 表 2 为 AGC 机组的参数. 其中, 光伏发电、风电和电动汽车不参与系统调频, 仅作负荷扰动处理.

考虑到众多新能源的间歇性和强随机性, 再通过引入随机负荷 (幅值 1000 kW, 周期 300 s) 信号模拟真实电网环境中所面临的不确定性, 进行 24 小时实时仿真, 验证  $DQ(\sigma, \lambda)$  的实际工程应用效果. 分别针对嵌入了  $DQ(\sigma, \lambda)$ 、 $PDW\omega LF-PhC(\lambda)$ 、 $Q(\sigma)$ 、 $Q(\lambda)$ 、 $Q$  等 5 种算法的控制器进行仿真. 图 9 为各控制器输出曲线 (方便效果对比, 仅截取前 2 000 s). 相较其他算法,  $DQ(\sigma, \lambda)$  控制器仿真曲线更加平滑、收敛速度更快; 图 10 为频率曲线, 上述 5 种算法最大频率偏差分别为 0.17 Hz、0.12 Hz、0.18 Hz、0.17 Hz、0.06 Hz 均满足实际工程要求, 可各控制器 A 区域 24 小时  $|\Delta f|$  分别为 0.0005 Hz、0.0013 Hz、0.0017 Hz、0.0027 Hz、

0.0045 Hz, 对比可知所提算法频率调整效果最优调整时间更短; 图 11 为区域联络线功率偏差曲线,  $P_{tie-max}$  (最大联络线交换功率偏差) 分别为 22 kW、108 kW、18 kW、27 kW、99 kW, 除 Q ( $\sigma$ ) 算法, 均

能在之后保持在 3 kW 的范围内. 此外, A 区域 24 小时的 CPS1 分别为 199.9807%、199.9338%、199.9201%、199.7749%、199.4645%, ACE 分别为 0.8673 kW、1.6622 kW、2.5792 kW、4.6971 kW、

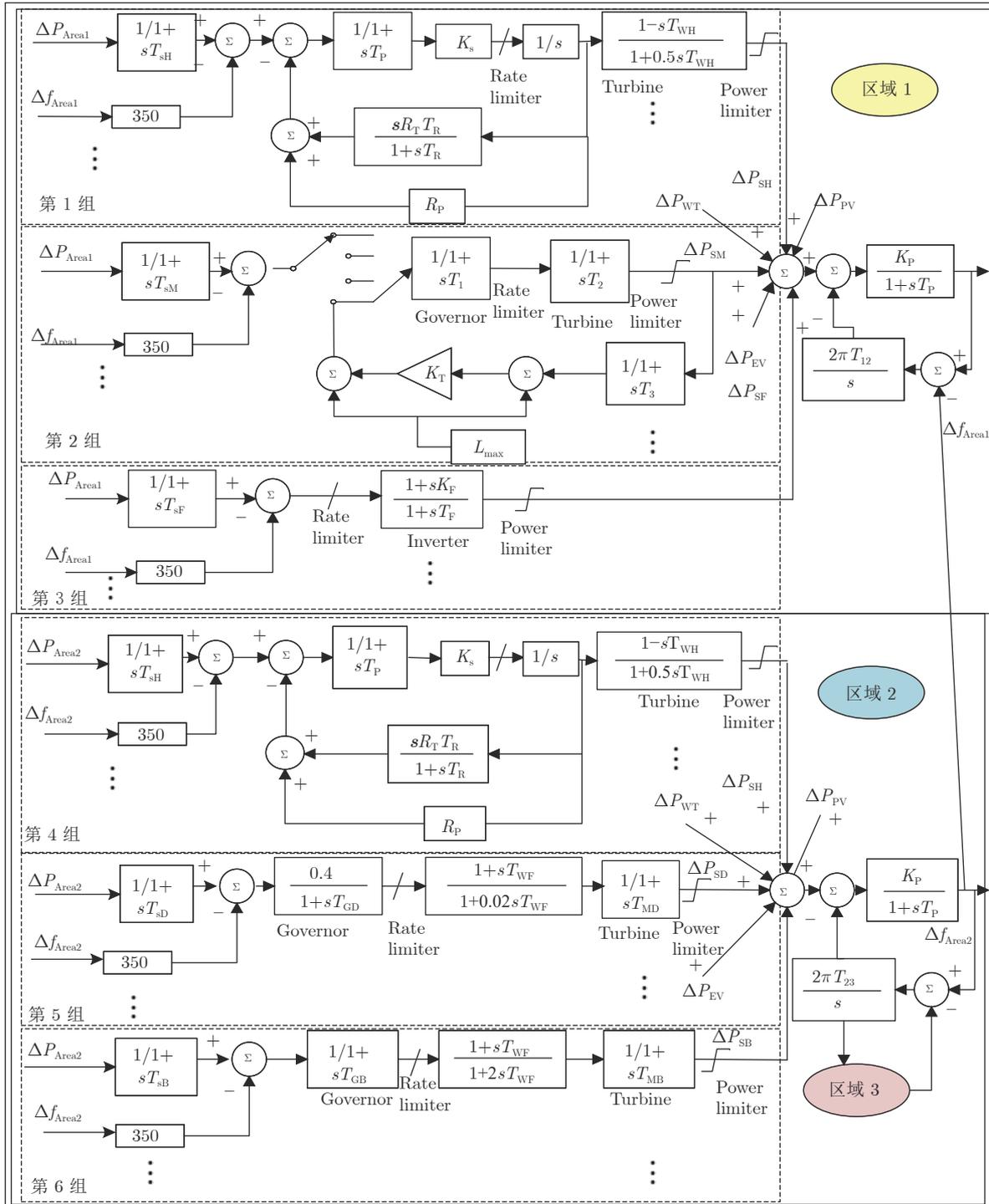


图 8 分布式 3 区域多能微网群协同 AGC 模型

Fig.8 Coordinated AGC model of a distributed three-area multi-energy microgrid group

表 1 模型传递函数的参数  
Table 1 Parameters of the model transfer function

机组	参数	数值
小水电机组	二次时延 $T_{SH}$	3
	伺机电动机时间常数 $T_P$	0.04
	伺机增益 $K_S$	5
	永态转差系数 $R_P$	1
	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
	闸门最大关闭率 $R_{maxclose}/(pu/s)$	0.16
	机组启动时间 $T_{WH}$	1
	二次时延 $T_{SB}$	10
生物发电机组	调速器的时间常数 $T_{GB}$	0.08
	蒸汽启动时间 $T_{WB}$	5
	机械启动时间 $T_{MB}$	0.3
微型燃气轮机机组	二次时延 $T_{SM}$	5
	燃油系统滞后时间常数 $T_1$	0.8
	燃油系统滞后时间常数 $T_2$	0.3
	负荷限制时间常数 $T_3$	3
	温度控制环路增益 $K_T$	1
	负荷限制 $L_{max}$	1.2
燃料电池机组	二次时延 $T_{SF}$	2
	调速器的时间常数 $T_F$	10.056
	逆变器增益 $K_F$	9.205
柴油发电储能机组	二次时延 $T_{SD}$	7
	调速器的时间常数 $T_{GD}$	2
	蒸汽启动时间 $T_{WF}$	1
	机械启动时间 $T_{MD}$	3

5.6927 kW. 上述指标进一步证明, 相较于其他传统算法, DQ ( $\sigma, \lambda$ ) 算法不仅均满足控制性能标准 (如第 3.1.1 节所述), 能实现各自区域内的电热功率平衡, 具有最优控制性能, 进而能够对分布式多区域多能微网群进行协同控制. 同时, 多区域协同控制能够使区域间能量互补, 有效地缓和储能设备的运行压力, 提高了多区域综合能源系统的灵活性和可靠性, 提高资源的利用率.

## 4 结论

为了对综合能源模式下的分布式多区域进行协同控制, 本文搭建了融入大量分布式能源的分布式多区域多能微网群协同的 AGC 模型, 并针对该模型提出了一种多智能体协同的 DQ ( $\sigma, \lambda$ ) 控制算法.

所提算法融入了资格迹, 不仅用于解决强化学习的时间信度分配问题, 而且“后向估计”机理提供了一个逼近最优值函数  $Q^*$  的渐进机制, 可提高

表 2 AGC 机组参数  
Table 2 AGC unit parameters

区域	类型	机组序号	$\Delta P_{in}^{max}$ (kW/s)	$\Delta P_{in}^{min}$ (kW/s)	$\Delta P_{in}^{rate+}$ (kW/s)	$\Delta P_{in}^{rate-}$ (kW/s)	
区域1和 区域3	小水电	G1	250	-250	15	-15	
		G2	250	-250	15	-15	
		G3	150	-150	8	-8	
		G4	150	-150	8	-8	
		G5	150	-150	8	-8	
		G6	100	-100	7	-7	
		G7	100	-100	7	-7	
	微型燃气 轮机	G8	100	-100	1.2	-1.2	
		G9	100	-100	1.2	-1.2	
		G10	150	-150	1.8	-1.8	
		G11	150	-150	1.8	-1.8	
		G12	200	-200	7	-7	
		G13	200	-200	7	-7	
		燃料电池	G14	150	-150	6	-6
			G15	150	-150	6	-6
区域2	小水电	G1	250	-250	15	-15	
		G2	250	-250	15	-15	
		G3	150	-150	8	-8	
		G4	150	-150	8	-8	
		G5	150	-150	8	-8	
		G6	100	-100	7	-7	
		G7	250	-250	2	-2	
	柴油发电 机储	G8	250	-250	2	-2	
		G9	120	-120	1	-1	
		G10	120	-120	1	-1	
	生物质能	G11	200	-200	3	-3	
		G12	200	-200	3	-3	
		G13	200	-200	3	-3	
		G14	200	-200	3	-3	

AGC 机组功率调节快速性; 同时为解决策略探索过程中动作值的高估, 所提算法在  $Q(\sigma, \lambda)$  的基础上采用去耦“动作选择”和“动作评估”相结合的双重学习.

通过对改进的 IEEE 标准两区域负荷频率控制模型以及分布式 3 区域多能微网 AGC 模型进行仿真, 结果显示, 与其他智能算法相比, 所提算法能提高收敛速度 93.92%~98.98%; 在能源出力不确定、负荷随机波动的情况下, DQ ( $\sigma, \lambda$ ) 仍能保持稳定的控制效果, 区域  $\Delta f$  降低 61.54%~88.89%、区域联络线功率偏差降低 18.51%~79.62%、CPS1 提高 0.023%~0.25%、ACE 降低 47.82%~84.76%, 能获得综合能源模式下分布式多区域协同.

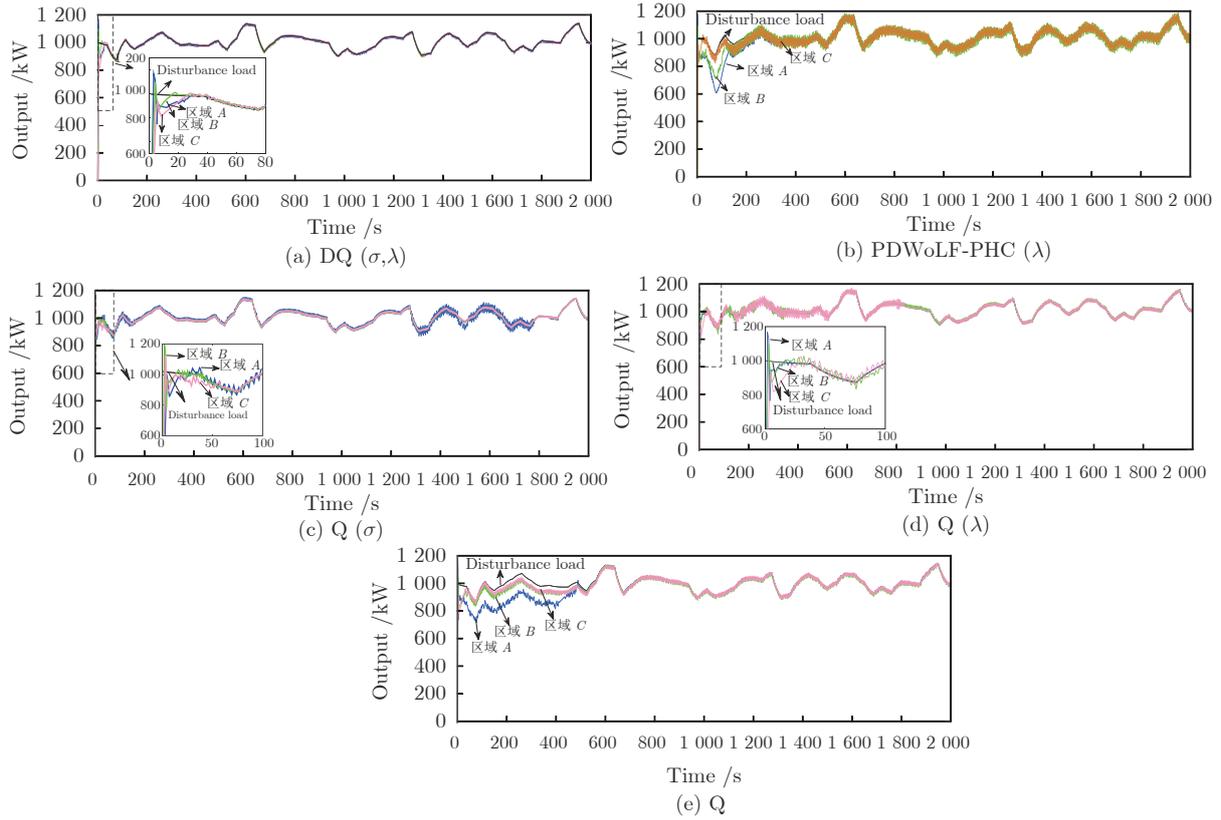


图 9 多算法输出效果

Fig.9 Multi algorithm output effect

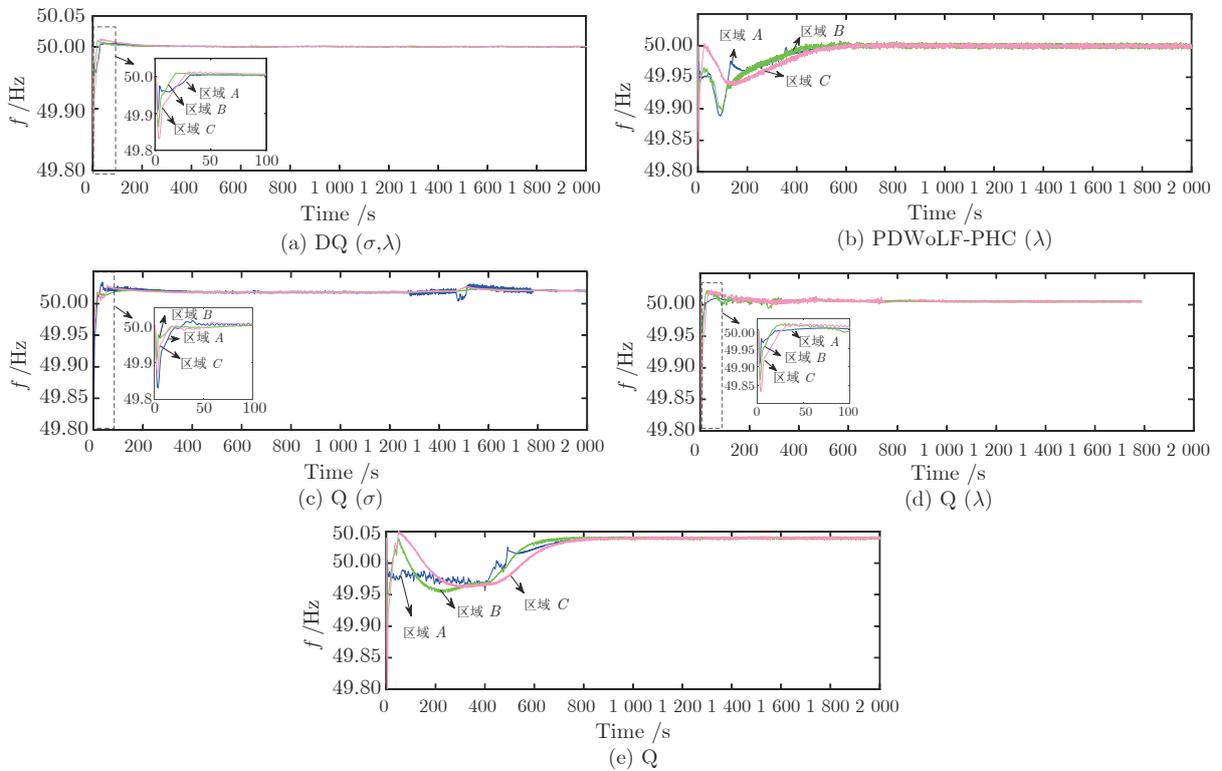


图 10 多算法频率曲线

Fig.10 Multi algorithm frequency curve

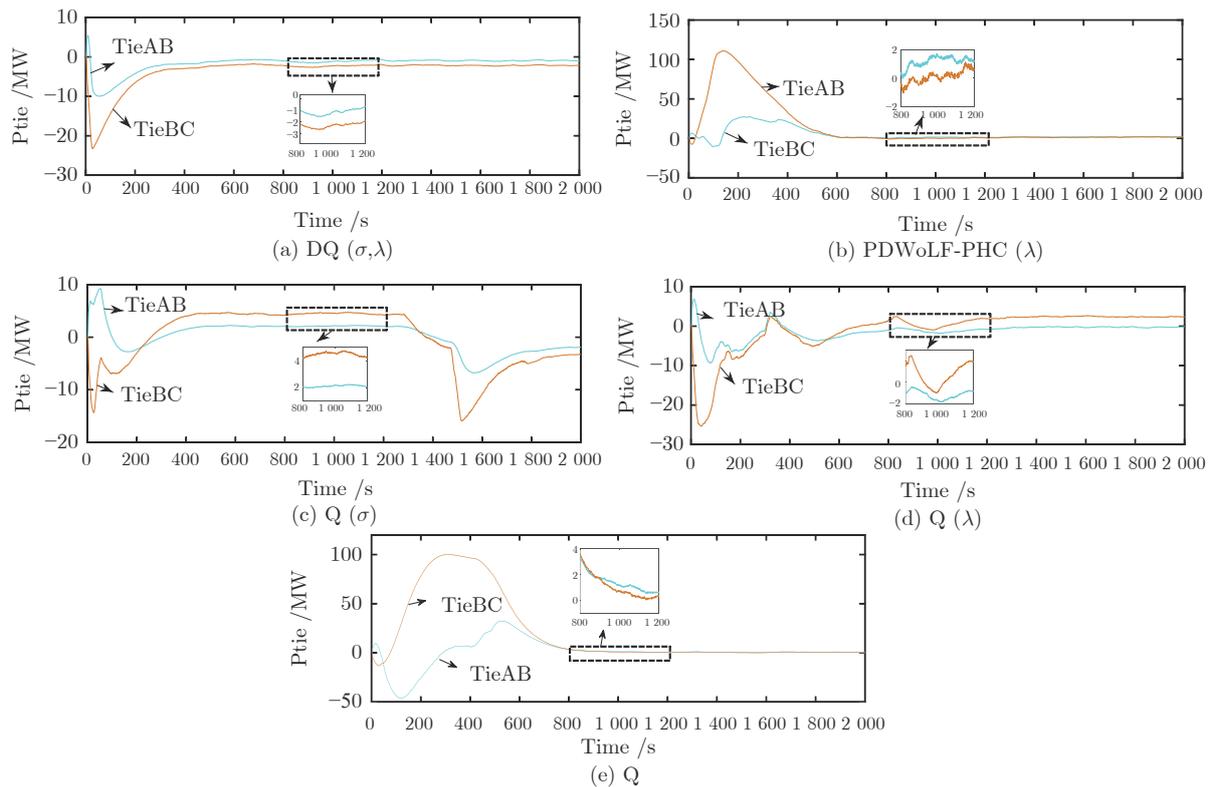


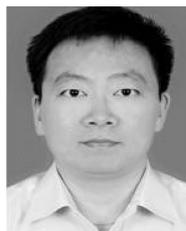
图 11 联络线交换功率偏差

Fig.11 Exchange power deviation of tie line

## References

- Meng L X, Savaghebi M, Andrad F, Vasquez J C, Guerrero J M, Graells M. Microgrid central controller development and hierarchical control implementation in the intelligent microgrid lab of Aalborg University. In: Proceedings of the 2015 IEEE Applied Power Electronics Conference and Exposition (APEC), Charlotte, NC, USA: IEEE, 2015. 2585-2592
- Brijesh P, Jiju K, Dhanesh P R, Joseph A. Microgrid for sustainable development of remote villages. In: Proceedings of the 2019 IEEE Region 10 Conference, Kochi, India: IEEE, 2019. 2433-2438
- Wang J, Cisse B M, Brown D, Crabb A. Development of a microgrid control system for a solar-plus-battery microgrid to support a critical facility. In: Proceedings of the 2017 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA: IEEE, 2017. 1-5
- Suyanto H, Irawati R. Study trends and challenges of the development of microgrids. In: Proceedings of the 6th IEEE International Conference on Advanced Logistics and Transport (IC-ALT), Bali, Indonesia: IEEE, 2017. 160-164
- Behera A, Panigrahi T K, Ray P K, Sahoo A K. A novel cascaded PID controller for automatic generation control analysis with renewable sources. *IEEE/CAA Journal of Automatica Sinica*, 2019, **6**(6): 1438-1451
- Jagatheesan K, Anand B, Samanta S, Dey N, Ashour A S, Balas V E. Design of a proportional-integral-derivative controller for an automatic generation control of multi-area power thermal systems using firefly algorithm. *IEEE/CAA Journal of Automatica Sinica*, 2019, **6**(2): 503-515
- Zhao Xi-Lin, Lin Zhen-Yu, Fu Bo, He Li, Xu Guang-Hui. Application of predictive optimization PID method to AGC of power system with windy power. *Journal of Power System and Automation*, 2019, **31**: 16-22 (赵熙临, 林震宇, 付波, 何莉, 徐光辉. 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)
- Xie Ping-Ping, Li Yin-Hong, Liu Xiao-Juan, Shi Dong-Yuan, Duan Xian-Zhong. Optimal PI/PID controller design of AGC based on social learning adaptive bacteria foraging algorithm for interconnected power grids. *Proceedings of the Chinese Society of Electrical Engineering*, 2016, **36**(20): 5440-5448 (谢平平, 李银红, 刘晓娟, 石东源, 段献忠. 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**(20): 5440-5448)
- Arya Y. A novel CFFOPID-FOPID controller for AGC performance enhancement of single and multi-area electric power systems. *ISA Transactions*, 2020, **100**: 126-135
- Xi L, Yu L, Xu Y C, Wang S X, Chen X. A novel multi-agent DDQN-AD method-based distributed strategy for automatic generation control of integrated energy systems. *IEEE Transactions on Sustainable Energy*, 2019, DOI: 10.1109/TSTE.2019.2958361
- Wu Xin, Shi Jun, Ma Wei-Zhe, Chen Jun-Bin. Automatic generation control of micro grid based on extreme Q-learning algorithm. *The Journal of New Industrialization*, 2019, **9**(4): 22-26 (吴新, 史军, 马伟哲, 陈俊斌. 基于极限 Q 学习算法的微电网自动发电控制. *新型工业化*, 2019, **9**(4): 22-26)
- Yu Tao, Liang Hai-Hua, Zhou Bin. Smart power generation control for microgrids islanded operation based on  $R(\lambda)$  learning. *Power System Protection and Control*, 2012, **40**(13): 7-13 (余涛, 梁海华, 周斌. 基于  $R(\lambda)$  学习的孤岛微电网智能发电控制. *电力系统保护与控制*, 2012, **40**(13): 7-13)
- Wu Li-Zhen, Lei Ai-Hu, Hao Xiao-Hong. Secondary control strategy of microgrid frequency of isolated island based on model predictive control. *Journal of Lanzhou University of Technology*, 2019, **45**(6): 99-107

- (吴丽珍, 雷艾虎, 郝晓弘. 基于模型预测控制的孤岛微电网频率二次控制策略. 兰州理工大学学报, 2019, **45**(6): 99–107)
- 14 Li Wen-hao. Decentralized Multi-Agent Reinforcement Learning Algorithm Research. [Master thesis]. East China Normal University, China, 2019.  
(李文浩. 去中心化多智能体强化学习算法研究 [硕士学位论文]. 华东师范大学, 中国, 2019.)
- 15 Qi Xiao. Research on Microgrid Load Frequency Control Strategy Based on Multi-Agent System and Active Disturbance Rejection Control Algorithm [Ph.D. dissertation]. North China Electric Power University, China, 2019.  
(慕晓. 基于多智能体系统及自抗扰控制理论的微网负荷频率控制策略研究 [博士学位论文]. 华北电力大学 (北京), 中国, 2019.)
- 16 Cao Qian. Consensus Algorithms Of Multi-Agent Systems And Its Application On Micro-Grid [Ph.D. dissertation]. University of Electronic Science and Technology of China, China, 2016.  
(曹倩. 多智能体系统一致性算法及其在微网中的应用 [博士学位论文]. 电子科技大学, 中国, 2016.)
- 17 Yi Nan. Design and Simulation of Microgrid Distributed Coordination Control System [Master thesis]. North China Electric Power University, China, 2014.  
(衣楠. 微网分布式协调控制系统设计及仿真实现 [硕士学位论文]. 华北电力大学, 中国, 2014.)
- 18 Li Nan-Fang. Research on Control Algorithms Based on Multi-agent Technology of Microgrid [Master thesis]. North China Electric Power University, China, 2011.  
(李楠芳. 基于多智能体技术的微电网控制算法的研究 [硕士学位论文]. 华北电力大学, 中国, 2011.)
- 19 Xi L, Li Y D, Huang Y H, Lu L, Chen J F. A novel automatic generation control method based on the ecological population cooperative control for the islanded smart grid. *Complexity*, 2018, **2018**: 1–17
- 20 Watkins C J C H. Learning from Delayed Rewards. [Ph.D. dissertation]. King's College, Cambridge, England, 1989.
- 21 De Asis K, Hernandez-Garcia J F, Holland G Z, Sutton R S. Multi-step reinforcement learning: A unifying algorithm. *AAAI*, 2018, arXiv: 1703.01327
- 22 Hasselt H V. Double Q-learning. *Neural Information Processing Systems 23*, Curran Associates, Inc. 2613–2621
- 23 Sutton R S. Learning to predict by the methods of temporal differences. *Machine Learning*, 1988, **3**(1): 9–44
- 24 Van Seijen H, Van Hasselt H, Whiteson S, Wiering M A. A theoretical and empirical analysis of expected sarsa. In: Proceedings of the 2009 IEEE Symposium Conference on Adaptive Dynamic Programming and Reinforcement Learning. 2009. 177–184
- 25 Jaleeli N, Vanslyck L S. NERC's new control performance standards. *IEEE Transactions on Power Systems*, 1999, **14**(3): 1091–1099
- 26 Zhang X S, Yu T, Pan Z N, Yang B, Bao T. Lifelong learning for complementary generation control of interconnected power grids with high-penetration renewables and EVs. *IEEE Transactions on Power Systems*, 2018, **33**(4): 4097–4110
- 27 Huang Ji-Yuan. Study on Optimal Allocation and Control Strategy Design of Battery Energy Storage System for Power Grid Frequency Regulation [Ph.D. dissertation]. Hunan University, China, 2015.  
(黄际元. 储能电池参与电网调频的优化配置及控制策略研究 [博士学位论文]. 湖南大学, 中国, 2015.)
- 28 Sun Q Y, Huang B N, Li D S, Ma D H, Zhang Y B. Optimal placement of energy storage devices in microgrids via structure preserving energy function. *IEEE Transactions on Industrial Informatics*, 2016, **12**(3): 1166–1179
- 29 Xu D, Wu Q, Zhou B, Li C, Bai L, Huang S. Distributed multi-energy operation of coupled electricity, heating and natural gas networks. *IEEE Transactions on Sustainable Energy*, 2019, DOI: 10.1109/TSTE.2019.2961432
- 30 Yu T, Zhou B, Chan K W, Chen L, Yang B. Stochastic optimal relaxed automatic generation control in non-Markov environment based on multi-step Q ( $\lambda$ ) learning. *IEEE Transactions on Power Systems*, 2011, **26** (3): 1272–1282
- 31 Sun Q Y, Han R K, Zhang H G, Zhou J G, Guerrero J M. A multi-agent-based consensus algorithm for distributed coordinated control of distributed generators in the energy internet. *IEEE Transactions on Smart Grid*, 2015, **6**(6): 3006–3019
- 32 Saha A K, Chowdhury S, Chowdhury S P, Crossley A. Modelling and simulation of microturbine in islanded and grid-connected mode as distributed energy resource. In: Proceedings of the 2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century. Pittsburgh, PA, USA: IEEE, 2008. 1–7
- 33 Zhang X S, Li Q, Yu T, Yang B. Consensus transfer Q-learning for decentralized generation command dispatch based on virtual generation tribe. *IEEE Transactions on Smart Grid*, 2018, **9**(3): 2152–2165



**席磊** 三峡大学副教授. 2016年于华南理工大学获得博士学位. 主要研究方向为电力系统运行与控制, 自动发电控制, 智能控制方法. 本文通信作者. E-mail: xilei2014@163.com

(**XI Lei** Associate professor at China Three Gorges University. He

received his Ph.D. degree from South China University of Technology in 2016. His research interest covers load frequency control, artificial intelligence techniques and automatic generation control. Corresponding author of this paper.)



**周礼鹏** 三峡大学硕士研究生. 主要研究方向为自动发电控制.

E-mail: zlp197@126.com

(**ZHOU Li-Peng** Master student at China Three Gorges University. His main research interest is automatic generation control.)