基于深度Q学习的强鲁棒性智能发电控制器设计

殷林飞,余 涛
(华南理工大学 电力学院,广东 广州 510640)

摘要:在现代互联大电网背景下,研究了多区域强鲁棒性的智能发电控制策略。在Q学习的架构下,将深度神经网络的预测机制作为强化学习的动作选择机制,提出了一种具有强鲁棒性的深度Q学习算法,设计了基于该算法的智能发电控制器。针对智能电网下的智能发电控制问题,在多智能体系统的框架下采用所提深度Q学习算法进行控制,并与传统的PID、Q学习和Q(λ)算法进行对比。在IEEE标准2区域和以南方电网4区域为背景的仿真模型(采用了23 328种不同模型参数)中进行数值仿真,仿真结果验证了所提深度Q学习算法的可行性和有效性,也验证了所设计控制器的强鲁棒性。

关键词:深度Q学习;智能发电控制;强鲁棒性;深度神经网络;多智能体系统
中图分类号:TM 761^{+}.2 **文献标识码**:A DOI:10.16081/j.issn.1006-6047.2018.05.002

0 引言

随着互联电网智能化的发展(即智能电网(smart grids)[1]),参与自动发电控制AGC(Automatic Generation Control)二次调频的机组在不断动态变化,从而逐渐发展了智能发电控制SGC(Smart Generation Control)技术[2]。与此同时,各种新能源和间歇性能源的接入,也给智能电网的控制问题带来了新的挑战,不仅外部扰动不断变化,而且系统内部参数也在不断变化。

对于SGC,依赖于模型的最优策略或算法不能应用于动态模型中,主要在电网环境方面(间歇性新能源的加入[3-4]、电动汽车的接入给电网带来了较大挑战)[5-6]、电力市场(供求关系、市场实时电价与控制区域之间的博弈)、运行方式(运行方式切换时容易引起频率振荡)、控制策略(不同区域的协调控制问题,要从系统的角度去协调控制,而不是单个区域的控制策略最优)和控制目标方面(同时满足控制性能、经济性和环保等多目标最优)存在问题[2]。针对控制策略问题,目前有强化学习、改进的强化学习(如Q(λ)算法[7]和R(λ)算法[8])、人工神经网络ANN(Artificial Neural Network)[9-12]等算法。虽然这些智能控制算法能应对不同类型的外部扰动,但是当系统内部参数变化时,智能控制算法需要学习的时间较长,因此有学者采用模型参数辨识的方法进行研究[13]。然而该参数辨识一般是应用简单模型建立的参数辨识,当模型复杂、不清楚各个环节的大致模型、无法获取系统有多少环节、有多个参数需要辨识时,该参数辨识方法则有待深入研究。

而在智能控制算法领域,机器学习ML(Machine Learning)近几年成为热点话题,特别是在谷歌公司的人工智能研究团队——深智(DeepMind)[14-15]开展的围棋大赛之后更是成为热点,如文献[15-16]详细介绍了深度学习的分类与发展,文献[14,17-18]分析了深度学习在围棋等游戏中的应用。ML中的ANN可做分类和预测,对其改进后发展了深度强化学习DRL(Deep Reinforcement Learning)[18-20]。不断发展的ML还能解决多智能体系统MAS(Multi-Agent System)的问题,即通过带有深度Q学习DQL(Deep Q Learning)算法的智能体可在不断更新的奖励中寻找最优的动作,从而在整个环境中不断地进行博弈[22]。

因此,本文将ML中的深度神经网络DNN(Deep Neural Network),融入ML中的Q学习算法框架中,利用训练后的DNN替换Q学习算法中的动作选择机制,提升算法对系统的认知能力,从而首次提出了一种全新架构的DQL算法;并利用所提算法设计智能发电控制器,在由多区域智能体构成的多智能体系统中应用,特别是进行各参数(如类型和干扰等外部扰动,汽轮机的3个关键参数[13,23]、可调容量、爬坡率等内部扰动)可变的大规模仿真。

1 智能发电控制器

1.1 SGC模型

针对高速发展的现代互联电网出现的电网环境方面的变化、电力市场的改革、运行方式的切换、控制策略的改变和控制目标的不同等问题,SGC需在非标称参数下具备最优的控制性能,且SGC具有分布式结构,每个区域利用各自的算法在互联的电网中追求各自的最优控制。在SGC模型中,区域i的SGC模型如图1所示。

与AGC模型不同的是,控制区域的联络线功率

收稿日期:2017-07-18;**修回日期**:2018-03-15
基金项目:国家重点基础研究发展计划(973计划)项目(2013CB228205);国家自然科学基金资助项目(51477055)
Project supported by the National Basic Research Program of China(973 Program)(2013CB228205) and the National Natural Science Foundation of China(51477055)

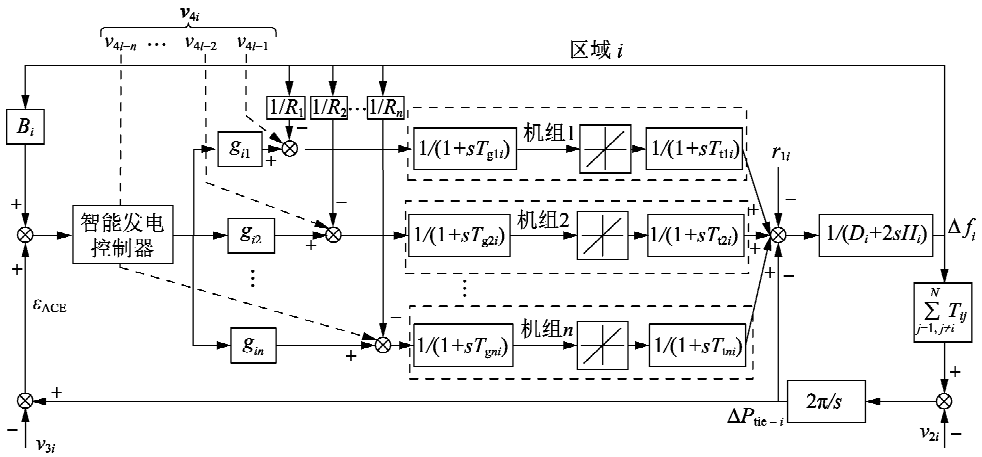


图 1　区域 i 的 SGC 模型

Fig.1　SGC model of Area i

变化不仅包含本地负荷扰动,而且包含本控制区域的基础负荷。图 1 中基础负荷 $P_{\mathrm{loc}i}$ 由与该地区签订的实时功率供需合同的发电机组来分担[2]。图 1 中,Δf_i 为区域 i 的频率偏差;B_i 为区域 i 的频率偏差系数;$\Delta P_{\mathrm{tie}-i}$ 为区域 i 的联络线总功率偏差;R_i 为区域 i 的下垂特性系数;H_i 为区域 i 的电力系统等值惯性常数;D_i 为区域 i 的电力系统等值阻尼系数;T_{gni} 为区域 i 第 n 台发电机组调速器的时间常数;T_{tni} 为区域 i 第 n 台发电机组的时间常数;N 为控制区域的个数;T_{ij} 为区域 i 和区域 j 之间的联络线同步系数。该分担由 $\boldsymbol{v}_{4i}$ 信号来实现。控制区域 i 的联络线功率变化 v_{2i} 为:

$$v_{2i}=\sum_{j=1,j\neq i}^{N}T_{ij}\Delta f_j \tag{1}$$

而该区域联络线有功功率计划值 v_{3i} 为:

$$v_{3i}=\sum_{j=1,j\neq i}^{N}\left(\sum_{k=1}^{n}g_{kj}\right)\Delta P_{\mathrm{loc}j}-\sum_{k=1}^{n}\left(\sum_{j=1,j\neq i}^{N}g_{jk}\right)\Delta P_{\mathrm{loc}i} \tag{2}$$

其中,g_{ki} 为第 k 台发电机机组在区域 i 的参与因子;$\Delta P_{\mathrm{loc}i}$、$\Delta P_{\mathrm{loc}j}$ 分别为区域 i、j 的有功功率差值。根据式(2)可得到任意控制区域的联络线功率偏差为:

$$\Delta P_{\mathrm{tie}-i,\mathrm{error}}=\Delta P_{\mathrm{tie}-i,\mathrm{actual}}-v_{3i} \tag{3}$$

其中,$\Delta P_{\mathrm{tie}-i,\mathrm{actual}}$ 为区域 i 联络线的实时功率。图 1 中的 $\boldsymbol{v}_{4i}$ 为外区域发电公司与本区域用电客户签订实时功率供需合同信息,即:

$$\boldsymbol{v}_{4i}=\begin{bmatrix}v_{4i-1} & v_{4i-2} & \cdots & v_{4i-n}\end{bmatrix} \tag{4}$$

其中,

$$\begin{cases}v_{4i-1}=\sum_{j=1}^{N}g_{1j}\Delta P_{\mathrm{loc}j}\\ \vdots\\ v_{4i-n}=\sum_{j=1}^{N}g_{nj}\Delta P_{\mathrm{loc}j}\end{cases} \tag{5}$$

在 SGC 中,发电机组 i 在 SGC 模型中发电总功率为:

$$\Delta P_{\mathrm{m}i}=\sum_{j=1}^{N}g_{ij}\Delta P_{\mathrm{loc}j} \tag{6}$$

1.2　智能发电控制器的控制目标

图 1 中的智能发电控制器必须控制区域的频率偏差 $|\Delta f|$ 尽量小,从而平衡各地区带来的功率误差。因此,智能发电控制器的目的为使频率偏差 $|\Delta f|$ 和区域功率误差 ACE(Area Control Error)均为 0。

为衡量智能发电控制器的控制性能,NERC 在 1997 年提出了统计学性能指标,即 CPS 指标。$\varepsilon_{\mathrm{ACE}}$ 则为该区域的功率控制误差(单位为 MW),Δf 为频率偏差(单位为 Hz)。$\varepsilon_{\mathrm{ACE}}$ 和 Δf 越小,则控制性能越优。

首先,定义 CPS1 指标为:

$$\delta_{\mathrm{CPS1}}=(2-\sigma_{\mathrm{CF1}})\times 100\% \tag{7}$$

$$\sigma_{\mathrm{CF1}}=\frac{\sum\varepsilon_{\mathrm{ACE,AVE-min}}\cdot\Delta F_{\mathrm{AVE-min}}}{-10B_i n_0\varepsilon_1^2}$$

其中,$\varepsilon_{\mathrm{ACE,AVE-min}}$ 为 1 min ACE 的平均值;B_i 为控制区域 i 的频率偏差系数(单位为 10 MW/Hz);n_0 为该统计时间内的分钟数;ε_1 为互联电网对全年每分钟频率平均偏差的均方根的控制目标值。

CPS2 指标定义为:

$$\delta_{\mathrm{CPS2}}=\left(1-\frac{T_{\mathrm{u}}}{T_{\mathrm{s}}-T_{\mathrm{n}}}\right)\times 100\% \tag{8}$$

其中,T_{u}、T_{s} 和 T_{n} 分别为考核期间不合格时段、总时段和非考核时段。T_{u} 为 ACE 每 10 min 的平均值大于 T_{10} 的考核时段数。CPS 指标的判断为:

$$\delta_{\mathrm{CPS}}=\begin{cases}\text{合格}\quad \delta_{\mathrm{CPS1}}\geqslant 200\%\text{或}\\ \qquad 100\%\leqslant\delta_{\mathrm{CPS1}}<200\%,\delta_{\mathrm{CPS2}}\text{合格}\\ \text{不合格}\quad \delta_{\mathrm{CPS1}}<100\%\text{或}\\ \qquad 100\%\leqslant\delta_{\mathrm{CPS1}}<200\%,\delta_{\mathrm{CPS2}}\text{不合格}\end{cases}$$

智能发电控制器从电网中采集 ε_{ACE} 和 Δf，并依据式(7)计算 δ_{CPS1} 指标，以 δ_{CPS1} 和 ε_{ACE} 作为输入，以发电机功率指令作为输出。

2 基于 DQL 的控制器

2.1 Q 学习算法

Q 学习算法作为“外控制”是不依赖于模型的属于马尔科夫决策过程 MDP(Markov Decision Process)的控制算法，它通过不断更新的奖励值来实现动态的最优的控制。Q 学习算法的核心是智能体与环境进行交互。对于智能体而言，从环境中获取到状态 s 和奖励值 r，然而事实是奖励值一般由人为设定，包含在智能体中，应为智能体的一部分。Q 学习算法中矩阵 $\boldsymbol{Q}$ 和矩阵 $\boldsymbol{P}$ 的更新方式为：

$$Q(s,a)\leftarrow Q(s,a)+\alpha(R(s,s',a)+\gamma\max_{a\in A}Q(s',a)-Q(s,a)) \tag{9}$$

$$P(s,a)\leftarrow\begin{cases}P(s,a)-P(1-P(s,a)) & a'=a\\ P(s,a)(1-\beta) & a'\neq a\end{cases} \tag{10}$$

其中，s 和 s' 分别为当前状态和下一时刻状态；β 为概率分布因子；概率矩阵 $P(s,a)$ 的初始值为 $1/|A|$，$|A|$ 为动作集中动作的数量，且概率矩阵的范围是 $P(s,a)\in[0,1]$；α 为 Q 学习算法的学习率；γ 为折扣因子；$R(s,s',a)$ 为奖励值，奖励值函数依据控制目标而定；a 和 a' 分别为当前时刻的动作和下一时刻的动作值。本文中 Q 学习算法的智能体的奖励函数为：

$$R(s,s',a)=\begin{cases}10 & \delta_{CPS1}\geqslant 200\%\\ -\varepsilon_{ACE}^2-1\,000\Delta f^2 & \delta_{CPS1}<200\%\end{cases} \tag{11}$$

在 Q 学习算法中，算法稳定性和收敛性有一定的随机性。在概率矩阵选择动作值时，若某动作的概率过大(存在“过学习”)，且其他动作概率很小，此时若未选择概率最大的动作，则会随机地从动作集中选择一个动作进行试错。这种试错会给 Q 学习的收敛速度带来影响。在试错少量的几个动作之后，能预测到在该情况下选择其他动作带来的影响，而此时 DNN 则能够实现此预测功能。

2.2 DNN

DNN 采用深层次的神经网络作为基础，将多个受限波尔兹曼机 RBM(Restricted Boltzmann Machine)堆叠。在训练 DNN 时，采用无监督的逐层贪心训练方法(逐层进行训练)。在离线训练完成之后，可采用有监督的学习对网络进行边训练边利用；再假定所有可见和隐含单元均为二值变量(只能取 0 或 1)，即 $i,j,v_i,h_j\in\{0,1\}$。基于能量定义的 RBM 系统的能量定义为：

$$E(\boldsymbol{v},\boldsymbol{h}\mid\boldsymbol{\theta})=-\sum_{i=1}^{n}a_iv_i-\sum_{j=1}^{m}b_jh_j-\sum_{i=1}^{n}\sum_{j=1}^{m}v_iW_{ij}h_j \tag{12}$$

其中，W_{ij} 为链接权重；a_i、b_j 分别为可见元 i 和隐元 j 的偏置。此时 $(\boldsymbol{v},\boldsymbol{h})$ 的联合概率分布为：

$$P(\boldsymbol{v},\boldsymbol{h}\mid\boldsymbol{\theta})=\frac{e^{-E(\boldsymbol{v},\boldsymbol{h}\mid\boldsymbol{\theta})}}{Z(\boldsymbol{\theta})} \tag{13}$$

其中，$Z(\boldsymbol{\theta})=\sum_{\boldsymbol{v},\boldsymbol{h}}e^{-E(\boldsymbol{v},\boldsymbol{h}\mid\boldsymbol{\theta})}$ 为归一化因子(配分函数)。RBM 的层间有连接、层内无连接的结构决定了各个隐元的激活条件是独立的。其激活概率为：

$$P(h_j=1\mid\boldsymbol{v},\boldsymbol{\theta})=\sigma(b_j+\sum_i v_iW_{ij}) \tag{14}$$

其中，$\sigma(x)=\frac{1}{1+e^{-x}}$ 为 sigmoid 激活函数。

各个可见元的激活概率为：

$$P(v_i=1\mid\boldsymbol{h},\boldsymbol{\theta})=\sigma(a_i+\sum_j h_jW_{ij}) \tag{15}$$

2.3 DQL 算法

为避免在某状态下多个动作对应的概率相同时 Q 学习算法的不断试错，加速算法的收敛性，在 Q 学习算法的框架下加入 DNN 进行下一时刻动作的预测。设计了 DQL 算法的框架如附录中图 A1 所示。

从图 A1 能看出，DQL 的框架和 Q 学习相似，通过 DNN 学习加速其对系统的预测能力，通过 DNN 对动作选择机制的替换，形成 DQL 算法。图 2 中展示：在 DNN 未被训练或预测输出的下一时刻状态不在“理想状态面”附近时，“训练切换开关”应置为“训练 DNN”档，其他情况应置为“训练结束”档。理想状态面则由以“当前时刻”状态为 x 轴，以动作集为 y 轴，以理想状态构成 z 轴，平行于 oxy 的平面构成，“动作选择”则为每次迭代过程中选择理想状态面附近对应的动作作为输出。当 DNN 无法预测或预测出的状态不在“理想状态面”附近时，智能体自动将“训练切换开关”置为“训练 DNN”档。对 DQL 算法的训练见 2.4 节。

2.4 基于 DQL 的控制器的训练与互博弈

对于 DQL 算法，样本获取极其关键。DQL 算法的样本来自于离线训练和在线微调 2 种方式。离线的静态训练需要在不同的状态 s 下执行不同的动作 a，从而获取下一时刻的状态 s'。

离线训练时，对于 Q 学习、Q(λ)和 DQL 算法的样本训练，可在某种内部参数情况下采用不同幅值的阶跃输入作为外部扰动进行算法样本训练，为获取不同的“当前时刻”的状态，输入不同的阶跃一段时间(本文算例中取 1 000 s)后，待系统稳定在某状态后选择不同的动作进行仿真，获取“下一时刻”的状态作为样本。

在线微调训练时，为加快算法收敛速度，可单独进行单个区域的算法训练。多区域的在线训练为多个基于 DQL 算法的智能体(或称为控制器)之间的

"互博弈"过程,多个基于 DQL 算法的智能体之间的博弈过程可分多次进行,其流程如图 2 所示。

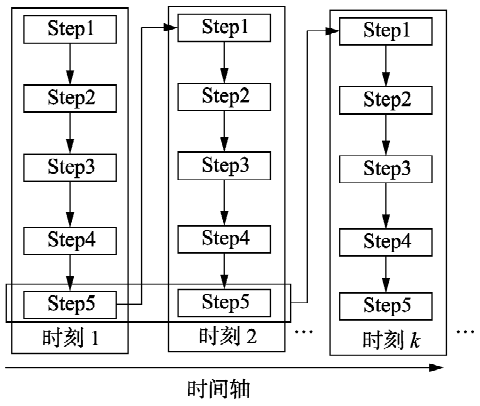


图 2　多 DQL 算法控制器的互博弈过程图

Fig.2　Game of multi-DQL controller

假定某智能互联大电网共有 4 个发电控制区域,从图 2 可看出多个基于 DQL 算法的控制器之间的互博弈过程为:

Step1　区域{2,3,4}固定选择单独训练得到的最优动作,区域{1}选择不同的动作进行训练;

Step2　区域{1}选择 Step1 训练得到的最优动作,区域{3,4}固定选择单独训练得到的最优动作,区域{2}选择不同的动作进行训练;

Step3　区域{1,2}选择[Step1,Step2]训练得到的最优动作,区域{4}固定选择单独训练得到的最优动作,区域{3}选择不同的动作进行训练;

Step4　区域{1,2,3}选择[Step1,Step2,Step3]训练得到的最优动作,区域{4}选择不同的动作进行训练;

Step5　区域{1,2,3,4}采用各自区域的 DQL 算法进行选择。

在训练"当前时刻"的样本时,在"当前时刻"之前的所有时刻的 4 个区域都用各自的 DQL 算法进行博弈。

2.5　DQL 算法的智能发电控制器设计

针对智能电网的 SGC 问题,设计了基于 DQL 算法的智能发电控制器,其结构如图 3 所示。

通过图 3 可以看出,以 DQL 算法为基础设计的区域 i 的智能发电控制器,将 ε_{ACE} 和 δ_{CPS1} 指标作为输入、机组出力 ΔP_{mi} 作为输出。该控制器通过 ε_{ACE} 和 δ_{CPS1} 确定所在状态,并更新矩阵 $\boldsymbol{Q}$。此后,若训练未结束,则更新概率矩阵 $\boldsymbol{P}$ 并训练 DNN,否则直接采用 DNN 进行预测并选择 Δf 最小对应的动作。针对智能发电控制问题,理想状态面可设定为 $\varepsilon=0.01$ Hz。

3　仿真算例

所有算例均在 CPU 为 i7－2760 2.4 GHz、内存为 8 GB 的电脑上运行,仿真软件的版本为 MATLAB

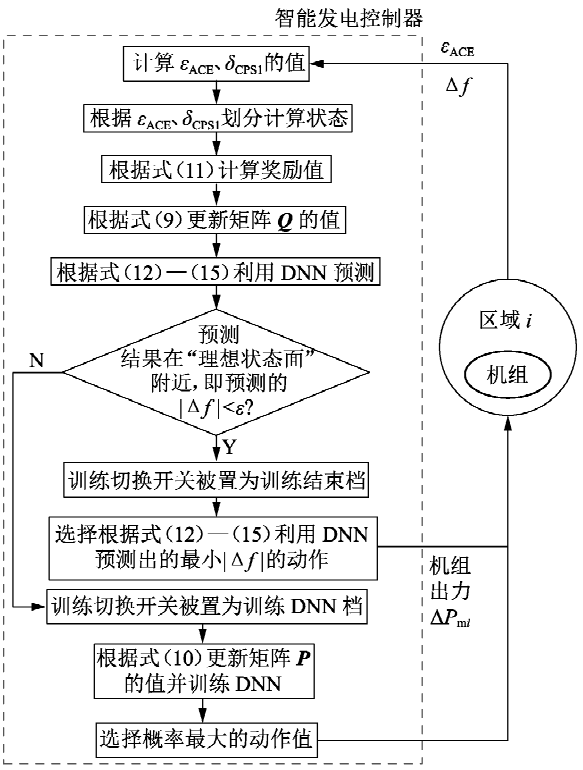


图 3　基于 DQL 算法的智能发电控制器

Fig.3　Smart generation controller based on DQL algorithm

2016b 9.1.0.441655。

3.1　IEEE 标准 2 区域模型

采用 IEEE 标准 2 区域模型作为算例进行仿真,扰动是周期为 1 200 s、幅值为 1 000 MW 的正弦扰动。模型中系统的基准容量为 5 000 MW,模型如附录中图 A2 所示,图中 $T_g=0.08$ s,$T_t=0.3$ s,$T_p=20$ s,$R=2.4$ Hz/p.u.,$K_p=120$ Hz/p.u.,$T_{12}=0.545$ s。

DQL、Q 学习和 Q(λ)算法中的矩阵 $\boldsymbol{Q}$ 和矩阵 $\boldsymbol{P}$ 的状态划分为 13 个,如表 1 所示。这些算法的动作区间取值为:{－50,－40,－30,－20,－10,0,10,20,30,40,50} MW。

表 1　2 区域模型 DQL、Q 学习和 Q(λ)算法的状态划分表

Table 1　State set of DQL, Q learning, and Q(λ) learning algorithms for two-area model

状态	ε_{ACE}或 δ_{CPS1}划分区间	状态	ε_{ACE}或 δ_{CPS1}划分区间
1	$\delta_{CPS1}>200\%$或 $\lvert\varepsilon_{ACE}\rvert<1$ MW	8	$-10\ \text{MW}\leqslant\varepsilon_{ACE}<-1$ MW
2	$1\ \text{MW}<\varepsilon_{ACE}\leqslant 10$ MW	9	$-20\ \text{MW}\leqslant\varepsilon_{ACE}<-10$ MW
3	$10\ \text{MW}<\varepsilon_{ACE}\leqslant 20$ MW	10	$-30\ \text{MW}\leqslant\varepsilon_{ACE}<-20$ MW
4	$20\ \text{MW}<\varepsilon_{ACE}\leqslant 30$ MW	11	$-40\ \text{MW}\leqslant\varepsilon_{ACE}<-30$ MW
5	$30\ \text{MW}<\varepsilon_{ACE}\leqslant 40$ MW	12	$-50\ \text{MW}\leqslant\varepsilon_{ACE}<-40$ MW
6	$40\ \text{MW}<\varepsilon_{ACE}\leqslant 50$ MW	13	$\varepsilon_{ACE}<-50$ MW
7	$\varepsilon_{ACE}>50$ MW		

分别采用 4 种算法进行仿真,将 DQL 算法和 PID、Q 学习和 Q(λ)算法进行对比,仿真结果如表 2 和图 4 所示。

表 2 2 区域仿真结果统计表

Table 2 Simulative results of two-area model

算法	δ_{CPS1}/%	δ_{CPS2}/%	ε_{ACE}/MW	Δf/Hz	δ_{CPS}/%
PID	198.204 7	100	47.112 03	0.010 437	100
QL	199.111 1	100	29.069 19	0.006 803	100
Q(λ)	199.604 7	100	19.959 19	0.004 785	100
DQL	199.716 3	100	18.804 04	0.004 394	100

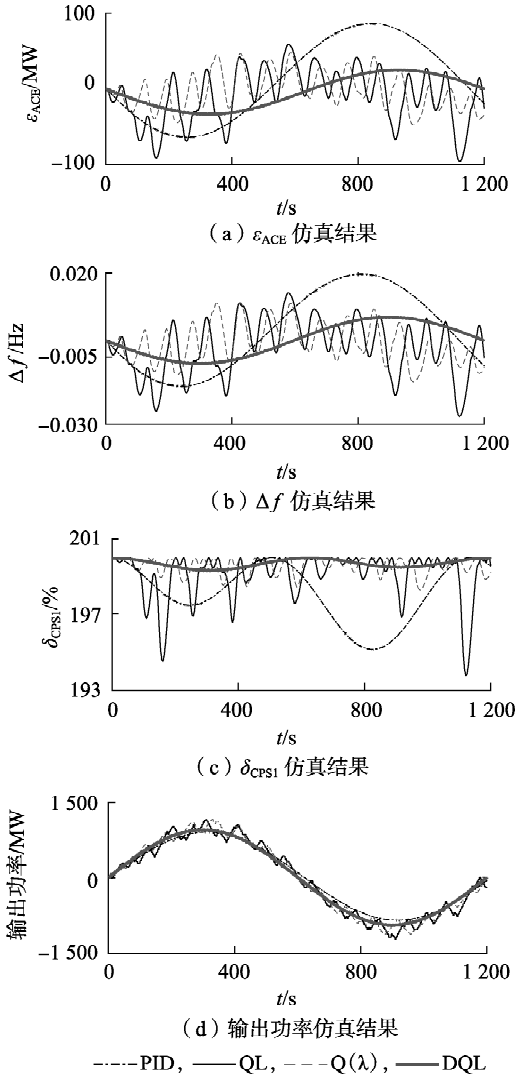


图 4 IEEE 标准 2 区域模型的仿真结果

Fig.4 Simulative results of IEEE standard two-area model

表 2 和图 4 中的 PID、QL、Q(λ)和 DQL 分别代表 PID、Q 学习、Q(λ)和 DQL 控制算法。

从表 2 可以看出,Q 学习、Q(λ)和 DQL 算法比 PID 算法的 ACE 和 Δf 小,且 DQL 最小。Q 学习、Q(λ)和 DQL 算法比 PID 算法的 ACE 分别小 38%、58%和 60%。Q 学习、Q(λ)和 DQL 算法比 PID 算法的 Δf 分别小 35%、54%和 58%。

且从图 4 也可以看出,DQL 算法的曲线比其他 3 种算法的曲线光滑、ACE 和 Δf 小、CPS 指标高。因此从仿真结果能看出 DQL 算法的效果优于其他 3 种算法。

3.2 以南方电网为背景的 4 区域模型

为验证所提 DQL 算法在复杂情况下的鲁棒性,在以南方电网为背景的 4 区域模型中进行大规模不同参数的数值仿真,在仿真中不仅变换外部扰动的类型和幅值,而且变换系统内部参数,来模拟系统本身的变化,如可调容量模拟丰水期和枯水期,再如汽轮机 3 个参数(T_{CH}、T_{RH}、T_{CO})、爬坡率 η_{GRC} 和二次调频时延参数 T_s 等参数的变换。所有参数可选取值如下:外部扰动波形有正弦波、方波、任意波扰动 3 种;风电接入扰动噪声取 0、10%、20%;T_s 取 8、20、30、35、60、120 s;η_{GRC} 取 3、5、8、10 p.u./min;可调容量取 1 000、500 MW;T_{CH} 取 0.2、0.25、0.3 s;T_{RH} 取 5、6、7、8、9、10 s;T_{CO} 取 0.3、0.4、0.5 s。该算例仿真模型如附录中图 A3 所示。3 种外部扰动在噪声为 0 情况下的波形如图 5 所示。

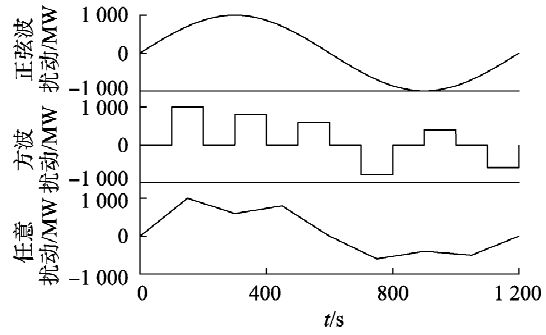


图 5 不同外部扰动曲线图

Fig.5 Curves of different external disturbances

可以看出,选择不同系统内部和外部参数时,共有 3×3×6×4×2×3×6×3=23 328 种组合,每种不同参数组合的模型需在线仿真 1 200 s,共需要 23 328×1 200=27 993 600 (s),即 324 d。每种组合需测试 4 种算法(PID、Q 学习、Q(λ)和 DQL 算法),共 324×4=1 296 (d)。

该算例中 DQL、Q 学习和 Q(λ)算法中的矩阵 $\boldsymbol{Q}$ 和矩阵 $\boldsymbol{P}$ 的状态也划分为 13 个,如表 3 所示。这些算法的动作取值为:{−500,−400,−300,−200,−100,0,100,200,300,400,500} MW。

表 3 4 区域模型 DQL、Q 学习和 Q(λ)算法的状态划分表

Table 3 State set of DQL, Q learning and Q(λ) learning algorithms for four-area model

状态	ε_{ACE}或 δ_{CPS1}划分区间	状态	ε_{ACE}或 δ_{CPS1}划分区间
1	δ_{CPS1}>200%或 $\lvert\varepsilon_{ACE}\rvert$<10 MW	8	−100 MW≤ε_{ACE}<−10 MW
2	10 MW<ε_{ACE}≤100 MW	9	−200 MW≤ε_{ACE}<−100 MW
3	100 MW<ε_{ACE}≤200 MW	10	−300 MW≤ε_{ACE}<−200 MW
4	200 MW<ε_{ACE}≤300 MW	11	−400 MW≤ε_{ACE}<−300 MW
5	300 MW<ε_{ACE}≤400 MW	12	−500 MW≤ε_{ACE}<−400 MW
6	400 MW<ε_{ACE}≤500 MW	13	ε_{ACE}<−500 MW
7	ε_{ACE}>500 MW		

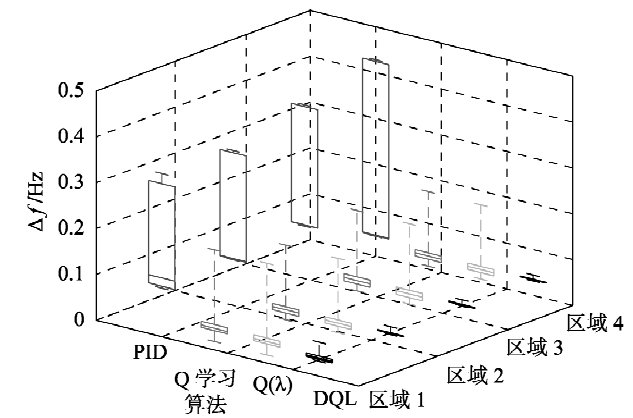
本算例中的 Q 学习、Q(λ)算法是在每种变参数的组合中单独训练的。而 DQL 算法在某一种参数

(参数如下:任意波扰动,噪声为 0,T_s = 30 s,η_{GRC} = 5 p.u./min,可调容量为 1 000 MW,T_{CH} = 0.25 s,T_{RH} = 8 s,T_{CO} = 0.3 s)下进行训练,在其他参数的情况下直接应用。

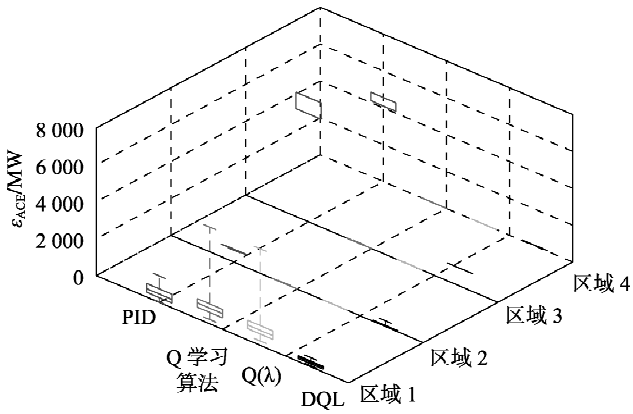
最后在模型中的 4 个区域都应用上述 4 种算法进行数值仿真,其结果统计如图 6—8 和表 4 所示(由于篇幅原因,只展示了不同扰动类型下其他不同参数组合的统计结果,其他不同参数组合的仿真结果与表 4 趋势一致)。

从表 4 可以看出:智能算法 Q 学习和 Q(λ)算法与所提 DQL 算法的 CPS 指标均比传统 PID 算法的 CPS 指标高 34.66%;DQL 算法比 Q 学习和 Q(λ)算法的 Δf 分别小 14.76%和 9.20%;在 δ_{CPS} 为 100%的情况下,DQL 算法比 Q 学习和 Q(λ)算法的 Δf 小。

从图 6 和图 7 可以看出:在系统参数和外部参数不断变化的过程中,PID 算法、Q 学习和Q(λ)算法得到的 Δf 和 ε_{ACE} 并非在每个区域都小,而 DQL 算法并非追求单一的 CPS 指标,而是满足综合 CPS 指标的情况下,尽量使得 Δf 最小;除区域 3 外,ε_{ACE} 和 Δf 以 DQL 算法最小,δ_{CPS} 以 DQL 算法为最大。



(a) 频率偏差 Δf 统计箱形图



(b) 区域控制误差 ε_{ACE} 统计箱形图

图 6　4 区域仿真结果统计箱形图

Fig.6　Statistics for four-area model(box chart)

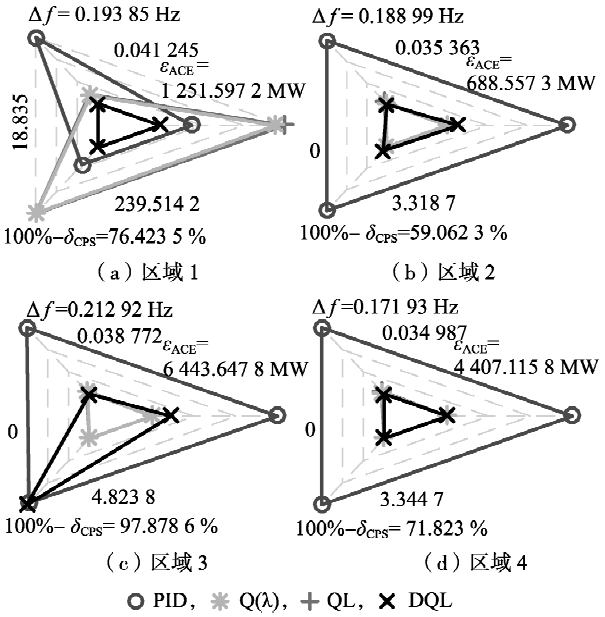


图 7　4 区域仿真结果统计蜘蛛网图

Fig.7　Statistics for four-area model(spider chart)

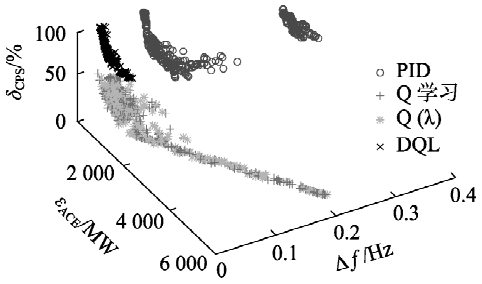


图 8　4 区域仿真结果状态分布图(区域 1)

Fig.8　Simulative results distribution of four-area model(Area 1)

表 4　4 区域仿真结果统计表(区域 4)

Table 4　Statistics for four-area model(Area 4)

扰动类型	算法	δ_{CPS1}/%	δ_{CPS2}/%	ε_{ACE}/MW	Δf/Hz	δ_{CPS}/%
方波	PID	198.696 8	100	6.998 723	0.013 752	66.36
	QL	199.887 2	100	3.436 284	0.037 363	100
	Q(λ)	199.863 1	100	3.389 566	0.036 728	100
	DQL	196.833 8	100	37.353 04	**0.035 156**	**100**
正弦波	PID	199.310 1	100	4.367 440	0.007 574	56.70
	QL	199.896 0	100	3.345 331	0.029 233	100
	Q(λ)	199.956 5	100	3.303 047	0.025 464	100
	DQL	197.348 7	100	37.036 55	**0.034 586**	**100**
任意波	PID	198.962 8	100	5.598 648	0.010 516	72.97
	QL	199.556 2	100	3.393 425	0.056 542	100
	Q(λ)	199.556 3	100	3.341 360	0.053 409	100
	DQL	197.015 1	100	37.225 23	**0.035 218**	**100**
总	PID	198.989 9	100	5.654 93	0.010 614	65.35
	QL	199.779 8	100	3.391 68	0.041 046	100
	Q(λ)	199.791 9	100	3.344 65	0.038 534	100
	DQL	197.065 8	100	37.204 94	**0.034 987**	**100**

图 8 的状态分布图是分别以 ε_{ACE}、Δf 和 δ_{CPS} 为 x、y 和 z 坐标轴绘出的区域 1 的性能分布图,可以看出,DQL 算法的控制性能优于其他 3 种算法控制性

能(DQL算法的δ_{CPS}高,且ε_{ACE}和Δf低)。

因此从该仿真结果可以看出,与PID算法、Q学习和Q(λ)算法相比,DQL算法的控制性能最优、算法更稳定,由其设计的控制器鲁棒性更强。

4 结论

针对智能电网中的SGC问题,提出了DQL算法,并设计了基于DQL算法的智能发电控制器,最后在IEEE标准2区域和以复杂南方电网为背景的大规模不同参数4区域模型(采用了23 328种不同模型参数)中进行数值仿真。所提DQL算法具有以下优点:

a. 与PID、Q学习和Q(λ)算法相比,所提DQL算法控制效果最优,验证了其解决SGC问题具有可行性和有效性;

b. 在大规模仿真实验中,基于所提DQL算法设计的智能发电控制器具有最强鲁棒性;

c. 所提DQL算法在多智能体系统中能够进行互博弈,从而探索最优控制过程。

在下一步工作中,将利用所提DQL算法设计更多电力系统控制器,如自动电压控制器、电力系统稳定控制器等。

附录见本刊网络版(http://www.epae.cn)。

参考文献:

[1] KEYHANI A,CHATTERJEE A. Automatic generation control structure for smart power grids[J]. IEEE Transac-tions on Smart Grid,2012,3(3):1310-1316.

[2] 王怀智. 智能发电控制的多目标优化策略及其均衡强化学习理论[D]. 广州:华南理工大学,2015.
WANG Huaizhi. Multi-objective strategy for smart generation control and equilibrium reinforcement learning theory[D]. Guangzhou: South China University of Technology,2015.

[3] 陈丽娟,姜宇轩,汪春. 改善电厂调频性能的储能策略研究和容量配置[J]. 电力自动化设备,2017,37(8):52-59.
CHEN Lijuan,JIANG Yuxuan,WANG Chun. Strategy and capacity of energy storage for improving AGC performance of power plant[J]. Electric Power Automation Equipment,2017,37(8):52-59.

[4] 李本新,韩学山,刘国静,等. 风电与储能系统互补下的火电机组组合[J]. 电力自动化设备,2017,37(7):32-37,54.
LI Benxing,HAN Xueshan,LIU Guojing,et al. Thermal unit commitment with complementary wind power and energy storage system[J]. Electric Power Automation Equipment,2017,37(7):32-37,54.

[5] 李清,张孝顺,余涛,等. 电动汽车充换电站参与电网AGC功率分配的成本一致性算法[J]. 电力自动化设备,2018,38(3):80-87,95.
LI Qing,ZHANG Xiaoshun,YU Tao,et al. Cost consensus algorithm of electric vehicle charging station participating in AGC power allocation of grid[J]. Electric Power Automation Equipment,2018,38(3):80-87,95.

[6] 程军. 风光互补智能控制系统的设计与实现[D]. 合肥:中国科学技术大学,2009.
CHENG Jun. Design and realization of hybrid wind/photovoltaic intelligent generation system[D]. Hefei: University of Science and Technology of China,2009.

[7] YU T,WANG H Z,ZHOU B,et al. Multiagent correlated equilibrium Q(λ) learning for coordinated smart generation control of interconnected power grids[J]. IEEE Transactions on Power Systems,2015,30(4):1669-1679.

[8] 余涛,梁海华,周斌. 基于R(λ)学习的孤岛微电网智能发电控制[J]. 电力系统保护与控制,2012,40(13):7-13.
YU Tao,LIANG Haihua,ZHOU Bin. Smart power generation control for microgrids islanded operation based on R(λ) learning[J]. Power System Protection and Control,2012,40(13):7-13.

[9] ZEYNELGIL H L,DEMIROREN A,SENGOR N S. The application of ANN technique to automatic generation control for multiarea power system[J]. International Journal of Electrical Power & Energy Systems,2002,24(5):345-354.

[10] CHEN D,KUMAR S,YORK M,et al. Smart Automatic Generation Control[C]// Power and Energy Society General Meeting. San Diego,California,USA:IEEE,2012:1-7.

[11] SAIKIA L C,MISHRA S,SINHA N,et al. Automatic generation control of a multi area hydrothermal system using reinforced learning neural network controller[J]. International Journal of Electrical Power & Energy Systems,2011,33(4):1101-1108.

[12] IMTHIAS T P,NAGENDRA P S,SASTRY P S. A neural network based reinforcement learning controller for automatic generation control[C]// National Power Systems Conference,NPSC2002. Hyderabad,India:Indian Institute of Technology,2002:161-165.

[13] 盛锴,江效龙,魏乐. 基于功率响应的汽轮机调节系统模型参数辨识方法研究[J]. 电力系统保护与控制,2016,44(12):100-107.
SHENG Kai,JIANG Xiaolong,WEI Le. Research on parameter identification of turbine governing system based on power response characteristics[J]. Power System Protection and Control,2016,44(12):100-107.

[14] MNIH V,KAVUKCUOGLU K,SILVER D,et al. Human-level control through deep reinforcement learning[J]. Nature,2015,518(7540):529-533.

[15] 赵冬斌,邵坤,朱圆恒,等. 深度强化学习综述:兼论计算机围棋的发展[J]. 控制理论与应用,2016,33(6):701-717.
ZHAO Dongbin,SHAO Kun,ZHU Yuanheng,et al. Review of deep reinforcement learning and discussions on the development of computer go[J]. Control Theory & Applications,2016,33(6):701-717.

[16] 尹宝才,王文通,王立春. 深度学习研究综述[J]. 北京工业大学学报,2015(1):48-59.
YIN Baocai,WANG Wentong,WANG Lichun. Review of deep learning[J]. Journal of Beijing University of Technology,2015(1):48-59.

[17] 陈兴国,俞扬. 强化学习及其在电脑围棋中的应用[J]. 自动化学报,2016,42(5):685-695.
CHEN Xingguo,YU Yang. Reinforcement learning and its application to the game of go[J]. Acta Automatica Sinica,2016,42(5):685-695.

[18] PENG X B,BERSETH G,MICHIEL V D P. Terrain-adaptive locomotion skills using deep reinforcement learning[J]. Acm Transactions on Graphics,2016,35(4):81.

[19] DENG Y,BAO F,KONG Y,et al. Deep direct reinforcement learning

for financial signal representation and trading[J]. IEEE Transactions on Neural Networks & Learning Systems, 2017, 28(3): 653-664.

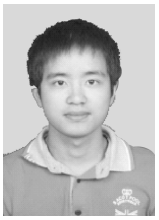
[20] LI L,LÜ Y,WANG F Y. Traffic signal timing via deep reinforcement learning[J]. IEEE/CAA Journal of Automatica Sinica, 2016, 3(3):247-254.

[21] JR G V D L C,DU Y,IRWIN J,et al. Initial progress in transfer for deep reinforcement learning algorithms[C]// Deep Reinforcement Learning:Frontiers and Challenges. New York,USA:[s.n.],2016: 1-6.

[22] 郑闻成. 基于 JADE 的多智能体动态博弈自动发电控制仿真平台研究[D]. 广州:华南理工大学,2014.
ZHENG Wencheng. Research on multiagent simulation platform for AGC Based on JADE[D]. Guangzhou: South China University of Technology,2014.

[23] 许天宁. 汽轮机电液调节系统模型参数辨识研究[D]. 吉林:吉林大学,2015.
XU Tianning. Research on model and parameter identification of the turbine DEH system[D]. Jilin:Jilin University,2015.

作者简介:


殷林飞

殷林飞(1990—),男,江西九江人,博士研究生,主要研究方向为智能发电控制(**E-mail**:yinlinfei@163.com);

余　涛(1974—),男,浙江宁波人,教授,博士研究生导师,通信作者,主要研究方向为智能发电控制、电能质量管理(**E-mail**:taoyu1@scut.edu.cn)。

Design of strong robust smart generation controller based on deep Q learning

YIN Linfei, YU Tao

(School of Electric Power, South China University of Technology, Guangzhou 510640, China)

Abstract: Under the background of modern interconnected large power grid, a smart generation control strategy with strong robustness in multi-areas is studied. In the framework of Q learning, taking the prediction mechanism of the deep neural network as the action selector of Q learning, a DQL(Deep Q Learning) algorithm with strong robustness is proposed, and on this basis, a smart generation controller is designed. The proposed DQL algorithm in the multi-agent system is applied for smart generation control in the smart interconnected power grid, and is compared with the traditional PID algorithm, Q learning algorithm and Q(λ) learning algorithm. The simulative results of IEEE standard two-area model and the four-area model based on China Southern Power Grid with 23 328 different parameters verify the feasibility and effectiveness of the proposed DQL algorithm and the strong robustness of the designed controller.

Key words: deep Q learning; smart generation control; strong robustness; deep neural network; multi-agent systems

附录

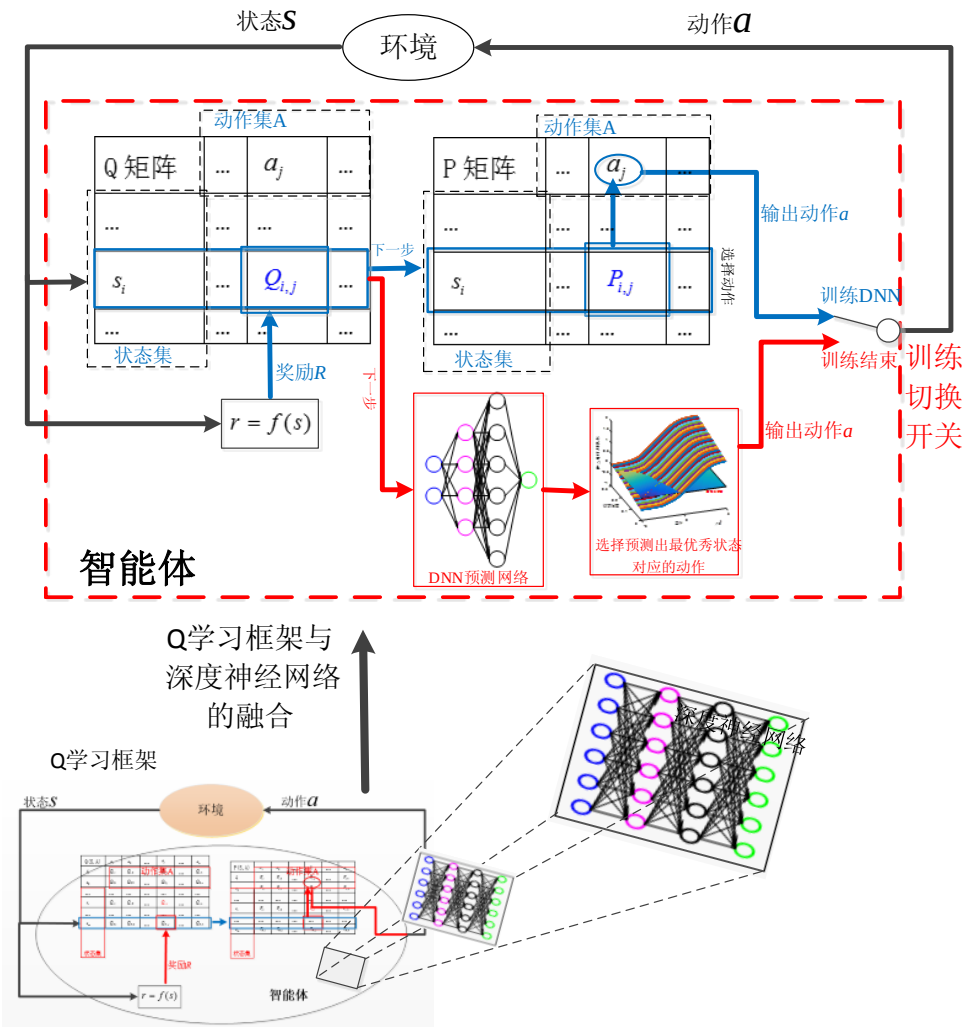


图 A1　DQL 算法框图

Fig.A1　Structure diagram of DQL algorithm

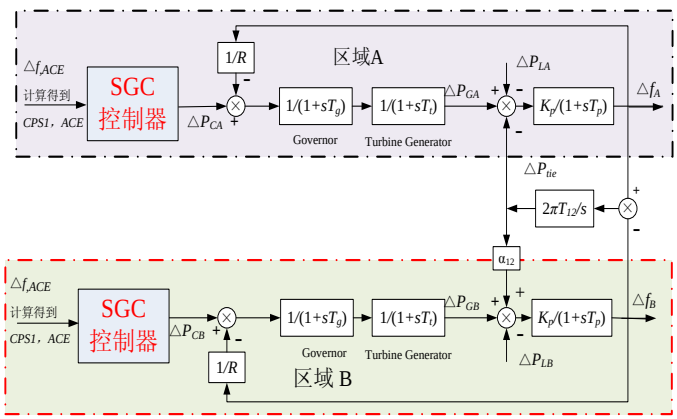


图 A2　IEEE 标准 2 区域仿真模型

Fig.A2　IEEE standard two-area simulation model

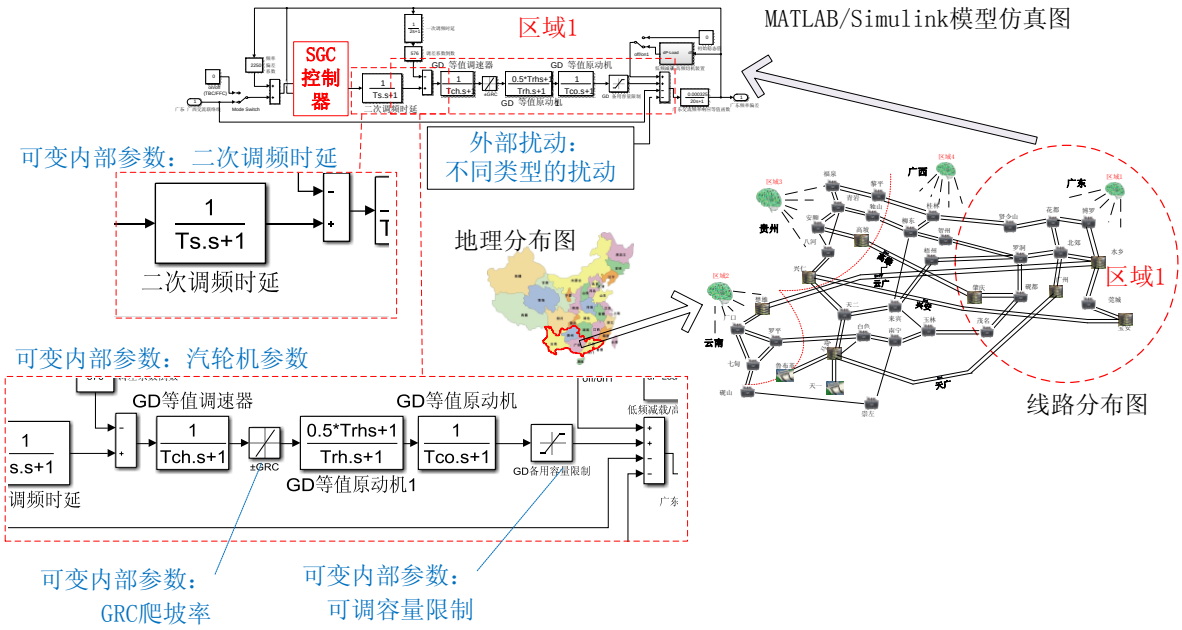


图 A3　以南方电网为背景的变参数 4 区域仿真模型

Fig.A3　Four-area model based on China Southern Power Grid