DOI:10.12158/j.2096-3203.2020.06.006

基于深度强化学习的电网自主控制与决策技术

王之伟[1,2]，陆晓[2]，刁瑞盛[1]，李海峰[2]，徐春雷[2]，段嘉俊[1]，张宁宇[3]，史迪[1]
（1. 全球能源互联网美国研究院，San Jose，CA 95134，美国；2. 国网江苏省电力有限公司，江苏 南京，210024；3. 国网江苏省电力有限公司电力科学研究院，江苏 南京，211103）

摘 要：高比例可再生能源的并网和电力电子设备的不断增加给电力系统运行与实时控制带来诸多挑战。人工智能技术的飞速发展为解决高维度、高非线性、高时变性优化控制和决策问题提供了新的思路。文中基于深度强化学习技术，提出了具有在线学习功能的电网自主优化控制和决策框架，即“电网脑”系统。该系统可通过离线和在线学习不断积累经验，从而在亚秒时间内（1 s 以内）根据电网实时量测数据给出调度控制指令及预期控制效果。该系统近期可用于辅助调度员决策，远期可为自动调度提供技术手段。为验证“电网脑”理论框架的可行性，文中以电网自主电压控制和联络线潮流控制为例，介绍了电力系统自主控制与决策方法及其实现流程，并通过数值实验验证了所提方法学习能力及其应用于电力系统自主控制与决策的可行性。

关键词：人工智能；电网脑；电网调度与控制；深度强化学习；亚秒级控制
中图分类号：TM854 **文献标志码**：A **文章编号**：2096-3203（2020）06-0034-10

0 引言

随着特高压交直流混联电网建设，高比例可再生能源持续接入，储能装置逐步应用以及电力市场规则、参与者行为的改变，电力系统电力电子化特征日趋显现，电网运行的不确定性、动态性和多元性显著增强。通常情况下，大电网的设计和运行理念是保证正常及故障工况下的安全稳定运行，制定包括电压、频率、线路潮流在内的多项安全指标[1—2]。但在某些突发重大故障时，若缺乏及时有效的在线控制，局部扰动可能会扩散导致连锁故障甚至大停电，如 2003 年 8 月北美东部大停电和 2011 年 9 月美国西南部大停电等事故[3]。因此，实时监测电网异常并采取快速、有效的调度控制措施对电网安全稳定运行至关重要。

数据采集与监控系统（supervisory control and data acquisition，SCADA）以及同步相量测量单元（phasor measurement unit，PMU）覆盖范围的扩大为大电网广域在线监测提供了有效途径。基于 PMU 数据驱动的高级应用也陆续被各级调控中心采用，应用范围包括：基于低秩矩阵的 PMU 数据质量分析与在线恢复[4—5]、基于 N+1 等效的电压稳定评估及预测[6—7]、基于扩展卡尔曼滤波的低频振荡监测和基于能量函数与机器学习的振源定位[8—10]、基于监督式机器学习的电压安全、暂态稳定评估等[11—13]。类似于许多其他广域量测系统（wide area measurement system，WAMS）在线应用，上述研究仅针对电网异常状况进行诊断，侧重于电网态势感知，难以直接给出快速且精准的实时控制策略。

近年来，先进人工智能技术，尤其是深度强化学习（deep reinforcement learning，DRL）技术不断进步，在多个领域（如 AlphaGo[14]，AlphaStar，无人驾驶，机器人[15]，负荷响应[16]等）成功应用，为电网智能自适应调控提供了启示。在电力系统优化和控制领域，文献[17]针对电力系统暂态稳定问题提出了基于强化学习的切机方案；文献[18—19]提出了基于强化学习的低频振荡抑制策略；文献[20]提出了交直流微网中基于强化学习的储能装置最优控制方法。然而，在电网在线调控领域，DRL 技术的研究与应用鲜见报道。

DRL 强大的学习与逻辑推演能力可为电力系统调度与控制提供更多的可能性，提升决策控制的速度。文中基于项目团队前期工作[21—22]，提出将 DRL 技术应用于电力系统调度控制的理论方法，并分析验证其可行性。提出了基于人工智能技术的数据驱动型电网智能调控技术框架，即“电网脑”系统。该系统由基于人工智能算法的智能体根据当前电网运行的态势感知结果，提供在线调控策略以及该策略预期效果。同时，“电网脑”系统在训练过程中，能够将传统控制策略转化为自身知识库，并根据实时量测数据进一步对控制策略进行优化和提升，以应对复杂多变的电网环境。“电网脑”系统的应用场景包括自主电压控制、频率控制、联络线潮流控制、最优潮流控制、电网网络拓扑实时优化、连锁故障防控和安全防御等，是对电网现有调控决

收稿日期：2020-05-05；修回日期：2020-06-16

策系统的补充与提升。文中以自主电压控制和联络线潮流控制为例,验证将人工智能技术应用于电网调控的方法论及其可行性。

文中论述了2种适用于不同控制动作集的DRL算法:深度Q网络(deep Q networks,DQN)和深度确定性策略梯度(deep deterministic policy gradient,DDPG)。分别以基于DRL的电网无功电压控制和联络线潮流控制为例,论述框架设计,奖励、状态和控制措施的制定以及算法实现流程,并在IEEE 14节点和美国伊利诺伊200节点系统中分别就自主电压控制与联络线潮流控制算法进行验证,并对2种算法的特点进行比较分析。

1 深度强化学习原理及核心算法

1.1 深度强化学习

人工智能是研究通过计算机程序模拟人类行为执行特定任务的技术科学。机器学习是人工智能的重要分支,可从复杂动态系统的大量观测数据中学习、训练并不断提升网络模型;使用时针对当前观测值立即给出相应输出结果。机器学习主要分为监督式学习、无监督式学习和强化学习(reinforcement learning,RL)3类[23]。

强化学习可有效解决复杂物理系统的控制和决策问题,图1为强化学习智能体与电力系统环境交互的过程。系统环境每执行一次智能体给出的动作(action),会返回新的系统状态(state)并计算相应的奖励值(reward);而智能体根据当前状态,以输出能够最大化奖励期望值的控制动作为目标,在与实际环境交互过程中不断学习并改进动作策略。文献[17]叙述了将强化学习描述成马尔可夫决策过程、值函数的定义以及几种最优策略的分析算法,此处不再赘述。

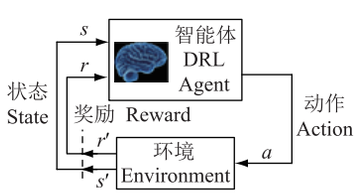

图1 强化学习智能体与环境交互流程

Fig.1 Interaction between RL agent and the environment

深度学习(deep learning,DL)提供通用化表征学习平台,通常使用多层非线性函数描述复杂物理系统的输入、输出关系。其优势在于可以从大量训练样本中自动搜寻有效样本特征来训练智能体并提升其性能,而无需提前人工指定。

DRL技术是DL与RL的结合,其中DL用于表征学习,而RL提供控制目标和策略。通过不断与环境交互迭代,DRL可进行自主学习,逐步提高决策、推理、预测能力,不断提升控制效果。此处的环境既可以是高精度电网仿真器,也可以是电网物理系统本身。文中重点介绍2种DRL核心算法,DQN和DDPG,分别适用于电力系统中控制变量为离散量和连续量的不同应用场景。

基于DRL技术提出数据驱动型电网智能调控技术框架,即"电网脑"系统,系统框架如图2所示。该系统由WAMS/SCADA数据驱动,实时采集、处理、分析大量在线数据并结合电网网络拓扑结构进行亚秒级状态估计,由智能体根据电网运行状态提供在线调控策略以及相应的预期控制效果。

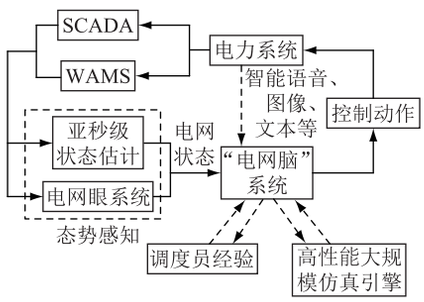

图2 基于DRL的智能调度系统框架

Fig.2 Architecture of the proposed intelligent autonomous control system based on DRL

1.2 DQN算法及适用范围

DQN使用深度神经网络对Q-学习算法进行有效扩展。传统Q-学习方法训练时需要使用Q表格记录每个训练样本的状态、动作和相应的Q值。为了解决Q-学习难以处理高维度状态和控制动作集的难题,在DQN方法中采用了神经网络模型实现Q函数值直接预测,从而避免使用Q表格。这使得Q函数可以利用连续的状态作为输入变量,大大提升了处理复杂问题的能力。在DQN方法中使用神经网络更新Q值的流程可描述为:

$$Q'_{(s,a)}=Q_{(s,a)}+\alpha(r+\gamma\max Q_{(s',a')}-Q_{(s,a)}) \quad (1)$$

式中:s, s'分别为系统当前与下一刻状态;a, a'分别为系统当前与下一刻控制动作;$\alpha\in[0,1]$为学习率;$\gamma\in[0,1]$为衰减率;r为奖惩函数值;$Q_{(s,a)}$为状态s下采取动作a的值函数的Q值。训练神经网络的目标是减小Q的预测值与真实值之间的误差,即$r+\gamma\max Q_{(s',a')}-Q_{(s,a)}$。为了提升DQN的学习效率,通常采用经验回放和定期修正目标网络Q值2种方法。首先,DQN分配内存专门存储历史经验并反复从中学习,进而更新策略。其次,为了避免不稳定而导致神经网络发散,采用2个独立的神

经网络模型：目标网络与评估网络。二者结构相同但参数不同。评估网络通过不断训练新样本来更新参数（更新较快），而目标网络从评估网络的参数中周期性更新（更新较慢）。该方法可有效提升DQN算法训练的稳定性。

1.3 DDPG算法及适用范围

DDPG是“演员-评论家”（actor-critic）方法和策略梯度（policy gradient）算法的有效融合[24]。该方法采用策略网络提供控制动作（起到“演员”作用），同时采用值函数网络来评估控制动作的效果（起到“评论家”作用）。类似于DQN方法，策略网络或值函数网络均可采用2个神经网络以不同速率更新其策略以提高训练效果。DDPG同样具有存储和回放历史经验的功能以提高训练性能，例如训练带有 N 个（$s_1,a_1,r_1,s_2,\cdots,s_N,a_N,r_N,s_{N+1}$）转移关系的随机样本子集（mini-batch）时，“演员”的策略网络可从最开始的 J 分布按照式（2）进行持续更新：

$$\nabla_{\theta^\mu}J\approx\frac{1}{N}\sum_i\nabla_a Q(s,a\mid\theta^Q)\mid_{s=s_i,a=\mu(s_i)}\nabla_{\theta^\mu}\mu(s\mid\theta^\mu)\mid_{s_i}\tag{2}$$

式中：控制动作可由“演员”策略函数 $a=\mu(s\mid\theta^\mu)$ 直接得出；θ^μ 为策略网络参数；θ^Q 为值函数网络参数。

定义 $y_i=r_i+\gamma\hat{Q}(s_{i+1},\hat{\mu}(s_{i+1}\mid\hat{\theta}^\mu)\mid\hat{\theta}^Q)$，其中（^）表示相应参数的估计值，则“评论家”的值函数网络可以按照式（3）来更新，以最小化偏差：

$$L=\frac{1}{N}\sum_i(y_i-Q(s_i,a_i\mid\theta^Q))^2\tag{3}$$

在DDPG算法中，目标网络的更新方式如式（4）所示，其中 τ 为一个较小的更新系数。

$$\begin{cases}\hat{\theta}^Q\leftarrow\tau\theta^Q+(1-\tau)\hat{\theta}^Q\\ \hat{\theta}^\mu\leftarrow\tau\theta^\mu+(1-\tau)\hat{\theta}^\mu\end{cases}\tag{4}$$

1.4 DQN与DDPG算法的比较及应用场景

DQN方法的实现相对容易，但需要针对每一个控制措施指定并计算相应的 Q 值，因此仅适用于控制/决策空间维度较低的离散型控制系统。DDPG算法可以有效处理大量、连续的控制变量，但相比DQN方法，其实现相对复杂，计算复杂度较高。另外，DDPG算法对于学习率、衰减率等超参数更为敏感，容易受到参数影响而达不到预期效果。为了避免过拟合，需要对DDPG算法中的超参数进行特殊处理：首先，使用较小的学习率和衰减率可有效避免瞬间过度学习情况发生；其次，随着训练的进行，学习率逐步降低，直至智能体完全掌握系统自主控制；再次，为了提高训练的有效性，训练过程可以采用优先经验回放等智能采样算法提高稳定性[24—25]；此外，考虑问题的复杂程度，深度神经网络的层数以及神经元的数量不宜过多。

DQN和DDPG算法各有优缺点，需要针对不同应用场景重新设计整体算法流程。例如，针对变压器变比或并联电容器开关这类离散控制变量问题，DQN算法更加有效；而针对发电机机端电压、有功/无功出力调节这类连续变量控制问题，选择DDPG更为合适。为了更好地对比2种算法，图3与图4详细阐述了基于DQN和DDPG算法的电力系统自主控制通用方法论以及智能体更新流程。

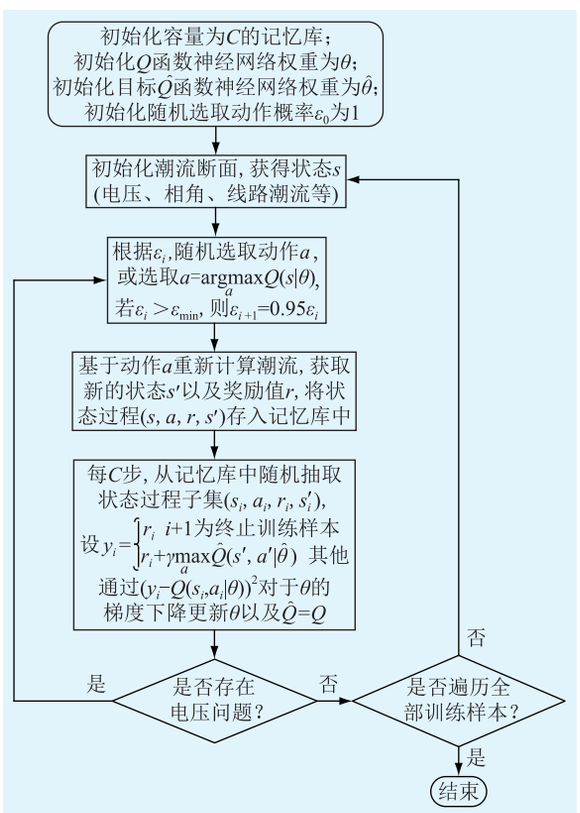

图3 基于DQN算法的电力系统自主控制流程

Fig.3 Flowchart for power grid autonomous control using DQN

图3中，在DQN智能体探索阶段，使用了衰减 ε-贪婪算法，即DQN在第 i 步控制迭代过程中，有 ε 的概率随机选取控制集中的任何控制措施，根据式（5）更新。

$$\varepsilon_{i+1}=\begin{cases}r_d\,\varepsilon_i & \varepsilon_i>\varepsilon_{\min}\\ \varepsilon_{\min} & \text{其他}\end{cases}\tag{5}$$

式中：r_d 为（0,1）区间的衰减常数。

如图4所示，在DDPG智能体探索阶段，更新策略时加入了随机衰减的噪音 ξ：

$$\mu'(s_i)=\mu(s_i\mid\theta^\mu)+\xi_i\tag{6}$$

式中：$\xi_{i+1}=r_d\,\xi_i$。

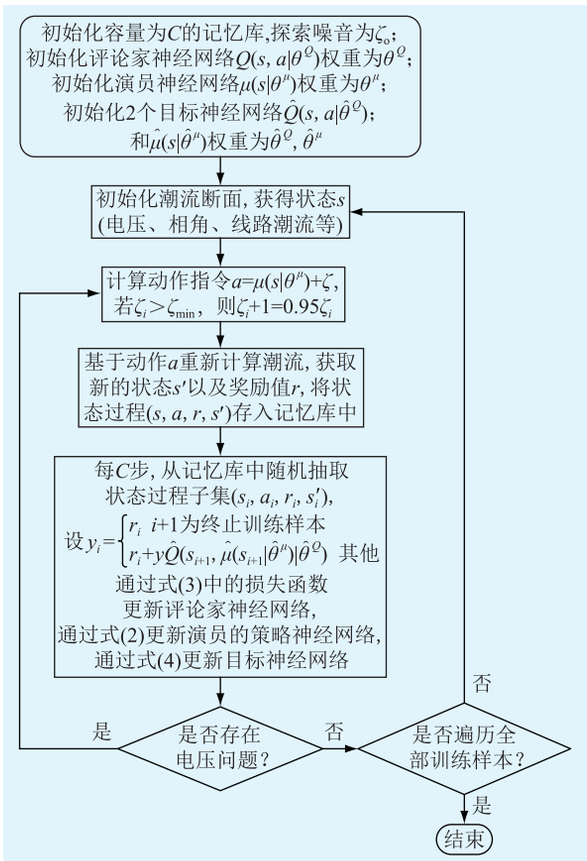

图 4 基于 DDPG 算法的电力系统自主控制流程

Fig.4 Flowchart for power grid autonomous control using DDPG

2 基于 DRL 的自主电压控制

2.1 控制目标和样本的定义

为了验证 DRL 在电力系统自主控制中应用的可行性,以自主电压控制为案例,通过智能体的不断学习和经验累积,提升其自主性和智能化水平,进而使电网各母线电压幅值在各种运行工况及扰动前后均能够维持在指定范围之内,一般为[0.95,1.05] p.u.。理想的控制目标(objective)是监测到电压越界后,智能体经过 1 次迭代直接给出最有效的控制策略解决系统电压问题。为了训练有效的 DRL 智能体,需要明确定义一个完整的训练样本(episode)以及相应的奖惩值、系统状态和控制动作集:

(1) 样本通过在线数据(SCADA、WAMS)或系统状态估计结果采集,可起始于稳态或准稳态时的任何工况。

(2) 若该工况无电压越界问题,则 DRL 智能体不需要提供控制措施,该逻辑可被设置为空的控制指令集。

(3) 若存在电压越界问题,DRL 智能体将被激发,通过迭代快速给出控制策略。迭代过程中,每组控制措施的效果可待电网进入新的准稳态后采集,或通过高精度电网潮流计算引擎获得。

(4) 若在规定的迭代次数内解决电压越界问题,则训练样本提前终止;若控制过程中潮流计算发散或迭代超过规定次数,该样本也将被强行终止。

2.2 奖励机制定义

每个有效样本的奖励机制定义如图 5 所示。一般可将系统节点电压幅值范围分为 3 个区域:

(1) 正常区域:[0.95,1.05] p.u.;

(2) 越限区域:[0.8,0.95) p.u.或(1.05,1.25] p.u.;

(3) 发散区域:[0,0.8) p.u.或 (1.25,+∞) p.u.。

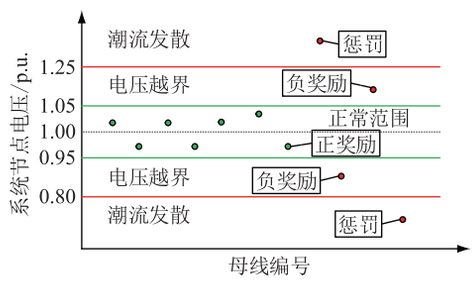

图 5 自主电压控制策略的奖励机制

Fig.5 Reward definition for autonomous voltage control

定义 V_j 为母线 j 的电压幅值,那么控制动作的第 i 次迭代奖励值可依据式(7)计算:

$$R_i=\begin{cases} 正奖励(+R_p) & \forall V_j\in[0.95,1.05]\text{p.u.} \\ 较小惩罚值(-R_n) & \exists V_j\notin[0.95,1.05]\text{p.u.} \\ 巨大惩罚值(-R_e) & 潮流发散 \end{cases} \tag{7}$$

通过式(7)可使 DRL 智能体朝着安全范围控制节点电压。为了得到更有效的控制策略(1 次迭代解决电压问题),每个训练样本的最终奖励值 R_f 可定义为一个完整样本内所有控制迭代所得奖励的平均值。

$$R_f=\sum_{i=1}^{n}R_i/n \tag{8}$$

式中:n 为完成一个训练样本所采取的控制迭代总次数。

采用该方式可有效促使 DRL 智能体以更少的控制迭代次数解决电压控制问题。需要指出的是,训练智能体时奖励值/函数的设定至关重要,对于不同控制问题和目标应采取不同的奖惩机制。例如,通过在奖惩函数中设计并加入成本函数,可以在系统运行条件约束下实现对不同目标(网损、线路剩余可用容量等)的优化[26—28]。

2.3 系统状态定义

样本的状态可从实时系统如 EMS(energy management system)或 WAMS 中获取,包括系统母线电压幅值、相角,线路有功、无功功率,发电机的出力,母线负荷等。为了有效应对不同类型量测值单位不同导致的灵敏度差异和误差,采用数据标准化的方式对所有类型的状态值进行统一处理[29]。假设在一个采样样本中有 m 组状态变量 $\boldsymbol{B}=\{x_1,\cdots,x_m\}$,首先计算样本的平均值 μ_{B}:

$$\mu_{\mathrm{B}}=\frac{1}{m}\sum_{i=1}^{m}x_i \tag{9}$$

该样本的变异系数定义为:

$$\sigma_{\mathrm{B}}^2=\frac{1}{m}\sum_{i=1}^{m}(x_i-\mu_{\mathrm{B}})^2 \tag{10}$$

则通过式(11)可计算出该样本的标准化值为:

$$x_i^*=\frac{x_i-\mu_{\mathrm{B}}}{\sqrt{\sigma_{\mathrm{B}}^2+\varepsilon}} \tag{11}$$

式中:ε 为提升数值稳定性的常数。

最后,样本可进一步增加比例系数和偏移量来提升样本训练效率,如式(12):

$$y_i=\gamma x_i^*+\beta \tag{12}$$

式中:γ 和 β 为可调整的样本系数。

2.4 控制动作集定义

为了简化问题,案例中以调节发电机端电压设定值为主要控制手段,阐述训练 DQN 和 DDPG 智能体实现自主电压控制的过程、经验和结论。

(1) 基于 DQN 方法的控制措施:每台发电机端电压设定值在离散控制集中选取,例如{0.95,0.975,1.0,1.025,1.05}p.u.。可选的控制空间可由所有参与调压的发电机组排列组合构成。

(2) 基于 DDPG 方法的控制措施:DDPG 算法可有效针对每个连续控制变量进行独立决策,因此智能体可根据每台发电机的物理参数决定相应控制量的取值范围,例如[0.9,1.05]p.u.。

2.5 基于 DRL 的自主电压调控实现流程

电压安全智能调度 DRL 智能体的训练流程如图 6 所示,分为 4 个主要步骤:

(1) 根据当前电网工况,随机改变负荷、发电机出力或添加故障,以模拟实际运行情况,如 $N-1$,求解电网潮流并检查所有节点电压幅值是否存在越界。

(2) 若存在越界情况,DRL 智能体将提供控制策略,反馈给电网环境(潮流计算引擎或 EMS 系统)进行验证。

(3) 电网环境执行控制指令,求解潮流获取控制动作后的电网状态并计算相应动作奖惩值。

(4) DRL 智能体从与环境交互中更新控制策略参数值并逐步提升控制性能。

需要指出的是,上述步骤为针对智能体的离线训练。在线训练或决策过程中,可以直接实施智能体的决策,不需要反馈给潮流计算引擎,这时可以直接通过电网实际量测值确定奖惩值;当然,也可以先反馈给潮流计算引擎或 EMS 系统进行验证,验证后再实施决策。

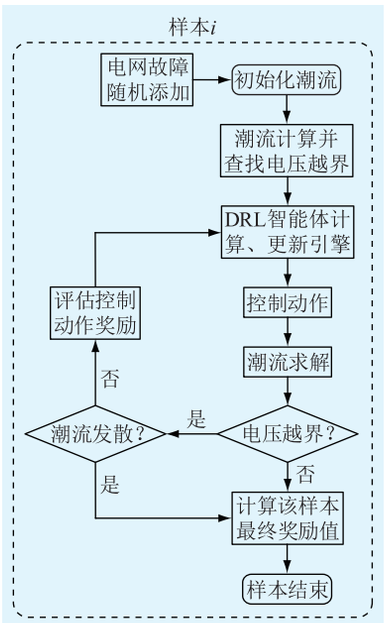

图 6 自主电压控制 DRL Agent 训练流程

Fig.6 Flowchart for training DRL Agent for autonomous voltage control

3 基于 DRL 的自主线路潮流控制

线路潮流是连续量,离散化控制效果不理想,因此采用 DDPG 算法较为适用。基于 DDPG 算法的自主潮流控制过程与电压控制整体类似,区别主要在于以下几点:

(1) 控制目标:电网所有或特定线路上的潮流在各种运行工况及扰动前后都在线路额定容量之内。

(2) 控制动作:除了平衡节点发电机外的机组有功出力在其有效控制范围 $[P_{\min},P_{\max}]$ 之内进行调节。

(3) 奖励机制:为了更好地关联奖惩和控制效果,鼓励低成本发电,同时惩罚线路和发电越限,奖惩的定义更新如式(13)。

$$R=\begin{cases}\dfrac{-C_{\mathrm{sys}}+E_1}{E_2} & \text{潮流正常}\\ \left(-\dfrac{D_{\mathrm{overflow}}}{E_3}-\dfrac{D_{\mathrm{pgen}}}{E_4}\right)/E_5 & \text{潮流越界}\end{cases} \tag{13}$$

式中:潮流正常指的是系统所有或特定线路的潮流在额定容量之内且各发电机有功输出在有效控制范围之内;C_{sys}为系统的发电成本,定义为式(14);$D_{overflow}$,D_{pgen}分别衡量系统线路潮流和发电机有功出力越限的程度,定义分别为式(15)和(16);E_1为在潮流正常情况下使奖励为正值的常量;$E_2 \sim E_5$的作用是使奖励的范围在[-1, 1]之间。

$$C_{sys} = \sum_{i=1}^{n} C_{gen,i}(P_{gen,i}) \tag{14}$$

$$D_{overflow} = \sum_{i=1}^{l} \{[\min(f_i - f_i^{min},0)]^2 + [\max(f_i - f_i^{max},0)]^2\} \tag{15}$$

$$D_{pgen} = \sum_{i=1}^{n} \{[\min(P_{gen,i} - P_{gen,i}^{min},0)]^2 + [\max(P_{gen,i} - P_{gen,i}^{max},0)]^2\} \tag{16}$$

式中:n,l分别为系统中发电机和线路总数;$C_{gen,i}(\cdot)$为发电机i的发电成本函数;$P_{gen,i}$为发电机i的有功出力;f_i,f_i^{min},f_i^{max}分别为线路i的实际潮流、最小容量和最大容量;$P_{gen,i}$,$P_{gen,i}^{min}$,$P_{gen,i}^{max}$分别为发电机i的实际、最小和最大有功出力。

4 算例分析及讨论

为了验证基于DRL算法的电力系统自主控制方法的有效性,对之前架构的自主电压控制以及联络线潮流控制模型分别进行测试。为了产生大量具有代表性的电网工况来训练DRL智能体,采用以下步骤产生训练样本:

(1) 了解并指定系统负荷变化的典型范围,如60%~140%。针对系统中每个负荷节点,随机改变负荷的有功值并维持负荷节点的功率因数不变。

(2) 当系统中所有负荷随机改变后,按一定比例调整发电机的有功值。该比例系数可根据发电机有功上限、可变化裕度、发电机类型等因素调整。

(3) 通过设定设备开断状态来反映设备检修、元件停运等电网拓扑结构的变化。

(4) 将所有变化反映在系统工况文件中,计算潮流,并保存有合理潮流解的工况。

采用加拿大Powertech Labs开发的商业潮流计算软件PSAT,按以上流程产生大量系统运行点(5万以上),并以PSS/E文本文件格式存储。

4.1 基于DRL的自主电压控制

在IEEE 14节点系统和美国伊利诺伊200节点系统[29]上对基于DQN和DDPG的自主电压控制算法进行有效性验证和详细性能测试。在14节点系统中,DRL智能体分别通过控制4台(DQN算例)、5台(DDPG算例)发电机电压来控制全系统节点电压;而在200节点系统中,DRL智能体通过同时调节38台发电机电压控制全网电压水平。在2个系统中,电压正常范围均定义为[0.95,1.05] p.u.。训练样本的电压越界情况统计如表1所示,其中绝大多数工况中存在母线电压越界。

表1 样本中工况的电压越限情况统计

Table 1 Summary of voltage violation for the created operating conditions

母线电压情况	越限母线数量	工况占比/%	
		IEEE 14节点	Illinois 200节点
低电压	>6	0.02	0
	5~6	0.47	0
	3~4	26.40	2.20
	1~2	19.41	19.36
无电压越界		8.25	10.56
过电压	1~2	0	5.53
	3~4	0.14	60.52
	5~6	7.97	1.83
	7~8	22.32	0
	>8	15.03	0

4.1.1 IEEE 14节点系统测试结果

首先训练DQN智能体对IEEE 14节点系统的50 000个运行工况进行自主电压控制;其中前40 000个用作训练模型,后1万个用作测试。训练采用双神经网络模型、归一化处理方法,并使用4台发电机的机端电压值构成了625组控制策略。DQN的控制效果如图7所示。随着训练次数增加,奖励值逐步提高。在测试集大多数样本中可以通过1次以内控制迭代解决电压问题,证实了控制策略的有效性。当增大控制动作空间(5台发电机同时控制,动作空间增大到$5^5 = 3\ 125$)时,发现许多测试样本需要2次以上迭代才能解决电压问题。由此可见,随着离散变量控制空间的增大,DQN算法的性能会有所降低。

DDPG方法可有效解决上述维数灾问题,其在相同样本下的训练和测试效果如图8所示。虽然训练开始时DDPG需要更多迭代次数对整个控制空间进行探索,但是在测试(实施)阶段,DDPG算法的控制效果优于DQN算法。

4.1.2 伊利诺伊200节点系统测试结果

为进一步测试基于DRL的自主电压控制方法的效果和鲁棒性,同时考虑到维数灾等因素,在伊利诺伊200节点系统[29]上测试了基于DDPG算法的自主电压控制策略,同时控制38台发电机母线电压的设定值,控制效果见图9、图10。

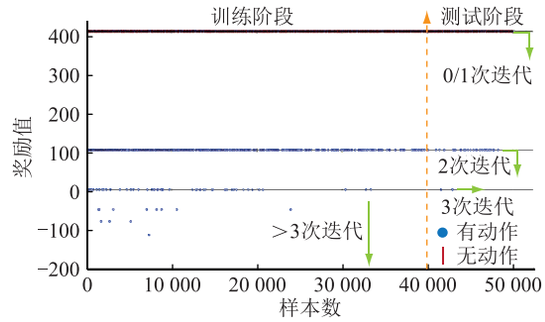

图 7 DQN Agent 在 IEEE 14 节点系统的测试效果

Fig.7 Performance of DQN Agent tested on the IEEE 14-bus system

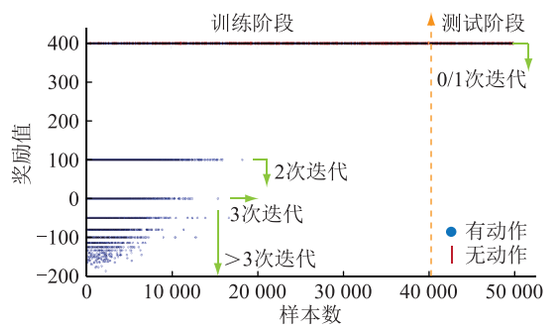

图 8 DDPG Agent 在 IEEE 14 节点系统的测试效果

Fig.8 Performance of DDPG Agent tested on the IEEE 14-bus system

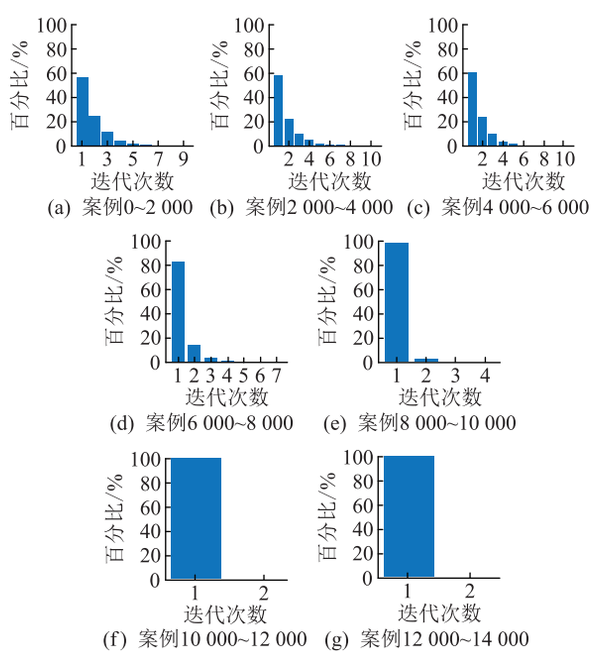

图 9 DDPG 智能体的迭代次数

Fig.9 Number of iterations the agent takes

图 9 为训练和测试样本集内智能体迭代次数统计;图 10 展示了奖励值随着样本数逐渐增大的过程,直至测试阶段所有样本均可在 1 次迭代以内解决电压问题。可见,在复杂度较高的 200 节点系统中,DDPG 算法同样可以快速、精准控制全系统电

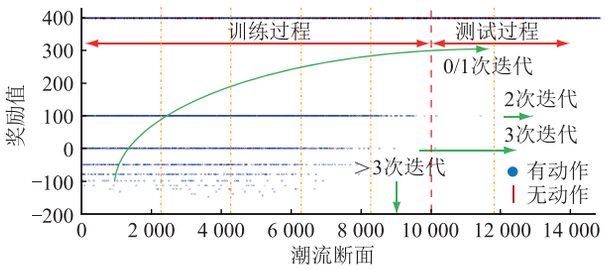

图 10 DDPG Agent 在伊利诺伊 200 节点系统的测试效果

Fig.10 Performance of DDPG Agent tested on the Illinois 200-bus system

压。针对美国伊利诺伊 200 节点系统,团队开发了测试和演示系统,实际测试性能和结果十分优异[30]。

4.1.3 不同奖惩值的性能比较

为了研究不同的奖惩机制能否激励智能体向着不同目标前进,案例中各节点电压都根据表 2 设定了不同的奖励值,以衡量其相对参考电压 V_{ref} 的偏离量,奖励机制的目标为使该偏离量最小。

表 2 各母线节点的奖惩设置

Table 2 Definition of reward for each bus

运行区间	母线节点电压 V_k/p.u.	奖励值 R_k	奖励值 R_k 变化区间
正常区间	$[V_{ref}, 1.05]$	$\frac{1.05 - V_k}{1.05 - V_{ref}}$	[0, 1]
正常区间	$[0.95, V_{ref}]$	$\frac{V_k - 0.95}{V_{ref} - 0.95}$	[0, 1]
电压越界	(1.05, 1.25]	$-\frac{V_k - V_{ref}}{1.25 - V_{ref}}$	[−1, −0.2]
电压越界	[0.8, 0.95)	$-\frac{V_{ref} - V_k}{V_{ref} - 0.85}$	[−1, −0.25]
潮流发散	(1.25, ∞)	−5	
潮流发散	[0, 0.8)	−5	

由图 11 可见,当 V_{ref} 不同(0.98 p.u. 或 1.0 p.u.)时,节点电压的平均量测会朝着指定的方向移动,进一步证明基于 DRL 的自动电压控制可以完成部分优化目标。

4.2 基于 DRL 的自主线路潮流控制

4.2.1 IEEE 14 节点系统测试结果

在 IEEE 14 节点系统上进行了基于 DDPG 的线路潮流控制的有效性验证,其中 30 000 个工况用于训练,10 000 个工况用于测试。在该系统中,DRL 智能体通过控制 4 台发电机的有功出力,使得该系统中 20 条线路的潮流均在额定容量范围内。由于该系统的原始参数中并没有线路容量,算例中采用了文献[31]中的容量限制。

该情况下的训练和测试结果如图 12 所示。经过 30 000 个样本的训练,DDPG 智能体在之后的

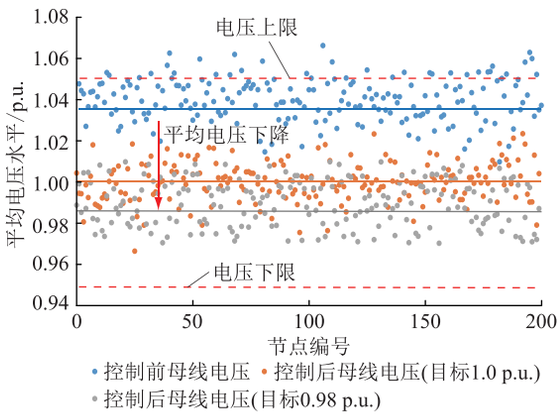

图 11 不同奖励机制对于 DRL 学习结果的影响

Fig.11 The illustration of the effect of reward on learning

10 000个测试样本中能够快速给出有效控制措施,使得所有线路潮流在额定容量范围内。其中智能体在 9 999 个(99.99%)样本中能够进行有效控制,其中在 99.7%的情况下能够在 1 步之内给出有效控制值,另 0.3%的情况下需要 2 步及以上迭代。

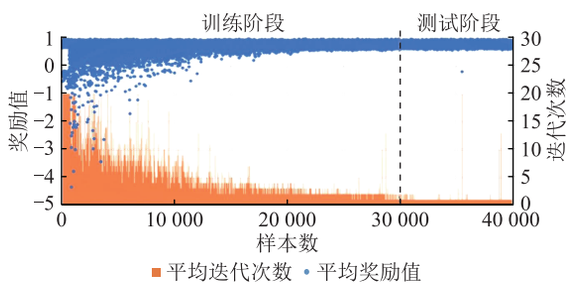

图 12 DDPG agent 在 IEEE 14 节点系统的训练和测试效果

Fig.12 Performance of DDPG Agent on the IEEE 14-bus system

4.2.2 伊利诺伊 200 节点系统测试结果

针对伊利诺伊 200 节点系统,通过随机负荷扰动产生 50 000 个样本,其中 40 000 个样本用于训练,10 000 个样本用于测试,训练和测试中采用文献[27]中的线路容量。训练和测试结果如图 13 所示。训练过程中为了提升智能体的稳定性,对智能体参数的更新在一定数量样本训练完成之后进行,即以 epoch 的方式对训练样本进行统计,每个 epoch 包含多个训练样本,因此训练阶段的结果显示的是每个 epoch 的平均迭代次数和奖励值。

经过 40 000 个样本的训练,智能体在之后的 10 000个测试样本中能快速给出有效控制措施,使所有线路潮流在额定容量范围内。其中智能体在所有 10 000 个(100.00%)测试样本中能够进行有效控制,在 99.98%的样本中能够在 1 步之内给出有效控制值,另 0.02%的样本中需要 2 步及以上迭代。

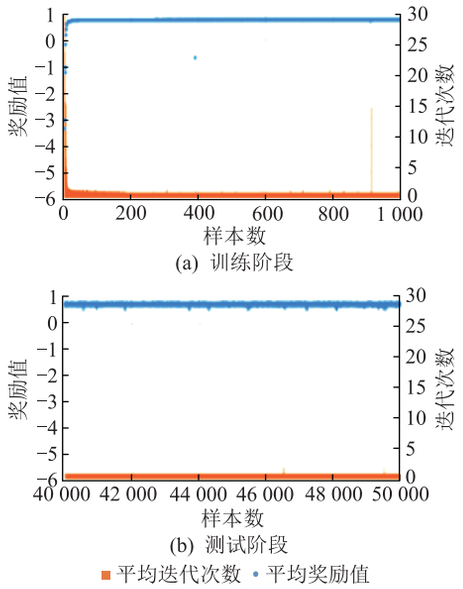

图 13 智能体在 200 节点系统的训练和测试效果

Fig.13 Performance of agent on the 200-bus system

5 结论及未来工作展望

文中提出了基于 DRL 算法的电网智能自主调控框架——“电网脑”系统,可针对调控问题在线快速提供解决方案,近期可用于辅助调度员决策,远期可为自动调度提供技术手段。其主要技术特性包括:在线动态挖掘数据与模型之间的关联,有效提取电网运行和控制的状态信息;具有亚秒级快速决策能力;可提供控制策略在线验证;适用于未覆盖的预想故障、非预设的运行方式等未知工况。基于该框架设计了自主电压控制与联络线潮流控制应用,并通过大量数值实验验证其有效性。

文中提出的技术框架与方法属于自主控制范畴,与电力系统现有自动控制系统,如自动电压控制、潮流控制等,有以下几点显著区别:(1) 电网现有自动控制系统的分析和设计建立在已知系统数学模型基础之上,而自主控制采取人类的思维方式,依靠数据和学习,建立逻辑模型,使用类似人脑的方法进行逻辑推演并实现控制目标;(2) 自动控制系统由设计者制定明确的规则,由控制器负责执行,系统输入、输出之间的逻辑关系是固定的,控制器本身不具备学习和进化能力,而自主控制可以学习被控对象和系统的规则,在控制器运行过程中通过自我组织、不断学习从而完成进化,同时获得非预知信息、积累控制经验并不断改善系统品质;(3) 自动控制的控制对象须为已知系统,当电力系统参数、拓扑、外界环境等发生变化时自动控制系统的模型和参数需要进行相应的调整和校核,而自主控

制的控制对象可以是已知或未知系统,其控制策略不仅可以应对外界干扰、复杂环境变化、参数改变等因素的影响,还可以消除模型化误差的影响。

未来研究工作包括:(1) 结合2种算法特点,设计并实现多智能体自主控制系统;(2) 进一步提升DRL算法训练速度和性能;(3) 将应用场景扩展至系统频率控制、电网经济性运行等方面;(4) 在奖惩机制设置过程中,考虑电网稳定性约束。

本文得到国网江苏省电力有限公司科技项目(SGTYHT/19-JS-215)资助,谨此致谢!

已申请美国发明专利:申请号 US 62/833,776,US 62/744,217。

参考文献:

[1] North American Electric Reliability Inc. Transmission system planning performance requirements[EB/OL]. [2020-04-29]. https://www.nerc.net/standardsreports/standardssummary.aspx.

[2] 国家能源局. GB 38755—2019 电力系统安全稳定导则[S]. 北京:中国标准出版社,2019.
National Energy Administration. GB 38755—2019 code on security and stability for power system[S]. Beijing: Standards Press of China,2019.

[3] U.S.-Canada Power System Outage Task Force. Final report on the August 14,2003 blackout in the United States and Canada [EB/OL]. [2020-04-29]. https://www3.epa.gov/region1/npdes/merrimackstation/pdfs/ar/AR-1165.pdf.

[4] LIAO M,SHI D,YU Z,et al. An alternating direction method of multipliers-based approach for PMU data recovery[J]. IEEE Transactions on Smart Grid,2019,10(4):4554-4565.

[5] LU X,SHI D,BIN Z,et al. PMU assisted power system parameter calibration at Jiangsu Electric Power Company[C]// IEEE PES General Meeting. Chicago,USA,2017:1-5.

[6] ZHANG X,SHI D,XIAO L,et al. Sensitivity based Thevenin index for voltage stability assessment considering N-1 contingency [C]//IEEE PES General Meeting. Portland,USA,2018:1-5.

[7] NIE Z,Zhang X,ZHAO X,et al. Adaptive online learning with momentum for contingency-based voltage stability assessment [C]//IEEE PES General Meeting. Atlanta,USA,2019:1-5.

[8] BIAN D,YU Z,DIAO R,et al. A real-time robust low-frequency oscillation detection and analysis (LFODA) system with innovative ensemble filtering[J]. CSEE Journal of Power and Energy Systems,2020,6(1):174-183.

[9] YU Z,SHI D,WANG Z,et al. Distributed estimation of oscillations in power systems:an extended Kalman filtering approach [J]. CSEE Journal of Power and Energy Systems,2019,5(2): 181-189.

[10] MENG Y,YU Z,SHI D,et al. Forced oscillation source location via multivariate time series classification[C]// IEEE PES T&D Conference and Exposition. Denver,USA,2018:1-5.

[11] DIAO R,VITTAL V,LOGIC N. Design of a real-time security assessment tool for situational awareness enhancement in modern power systems[J]. IEEE Transactions on Power Systems,2010,25(2):957-965.

[12] DIAO R,SUN K,VITTAL V,et al. Decision tree based online voltage security assessment using PMU measurements[J]. IEEE Transactions on Power Systems,2009,24(2):832-839.

[13] GENC I,DIAO R,VITTAL V,et al. Decision tree-based preventive and corrective control applications for dynamic security enhancement in power systems[J]. IEEE Transactions on Power Systems,2010,25(3):1611-1619.

[14] SILVER D,SCHRITTWIESER J. Mastering the game of Go without human knowledge[J]. Nature-International Journal of Science,2017,550:354-359.

[15] KOBER J,BAGNELL J,PETERS J. Reinforcement learning in robotics:a survey[J]. International Journal of Robotics Research,2013,32(11):1238-1278.

[16] BIAN D,PIPATTANSOMPORN M,RAHMAN S. A human expert-based approach to electrical peak demand management [J]. IEEE Transactions on Power Delivery,2015,30(3): 1119-1127.

[17] 刘威,张东霞,王新迎,等. 基于深度强化学习的电网紧急控制策略研究[J]. 中国电机工程学报,2018,38(1): 109-119.
LIU Wei,ZHANG Dongxia,WANG Xinying,et al. A decision making strategy for generating unit tripping under emergency circumstances based on deep reinforcement learning[J]. Proceedings of the CSEE,2018,38(1):109-119.

[18] DUAN J,XU H,LIU W. Q-learning based damping control of wide-area power systems under cyber uncertainties[J]. IEEE Transactions on Smart Grid,2018,9(2):6408-6418.

[19] HASHMY Y,YU Z,SHI D,et al. Wide-area measurement system-based low frequency oscillation damping control through reinforcement learning[J]. IEEE Transactions on Smart Grid, 2020,Early Access.

[20] DUAN J,YI Z,SHI D,et al. Reinforcement-learning-based optimal control for hybrid energy storage systems in hybrid AC/DC microgrids[J]. IEEE Transactions on Industrial Informatics,2019,15(9):5355-5364.

[21] DIAO R,WANG Z,SHI D,et al. Autonomous voltage control for grid operation using deep reinforcement learning[C]// IEEE PES General Meeting. Atlanta,USA,2019:1-5.

[22] DUAN J,SHI D,DIAO R,et al. Deep-reinforcement-learning-based autonomous voltage control for power grid operations[J]. IEEE Transactions on Power Systems,2019,35(15):814-817.

[23] Artificial Intelligence Industry. An overview by segment [EB/OL]. [2020-04-29]. https://www.techemergence.com/artificial-intelligence-industry-an-overview-by-segment/.

[24] LILLICRAP P,HUNT J,PRITZEL A,et al. Continuous control with deep reinforcement learning[C]//International Conference on Learning Representations. San Diego,USA,2015:1-14.

[25] DBADI M,BARHAM P,CHEN J,et al. Tensorflow:a system for large-scale machine learning[C]//Symposium on Opera-

ting Systems Design and Implementation. Savannah, USA, 2016:1-21.

[26] TU L, DUAN J, ZHANG B, et al. AI-based autonomous line flow control via topology adjustment for maximizing time-series ATCs[C]//IEEE PES General Meeting. USA, 2020:1-5.

[27] ZHANG B, LU X, DIAO R, et al. Real-time autonomous line flow control using proximal policy optimization[C]//IEEE PES General Meeting. USA, 2020:1-5.

[28] DUAN J, LI H, ZHANG X, et al. A deep reinforcement learning based approach for optimal active power dispatch[C]//IEEE Sustainable Power and Energy Conference. Beijing, China, 2019:1-6.

[29] XU T, BIRCHFIELD A, OVERBYE J. Modeling, tuning and validating system dynamics in synthetic electric grids[J]. IEEE Transactions on Power Systems, 2018, 23(6): 6501-6509.

[30] Grid mind demo[EB/OL]. [2020-04-29]. https://geirina.net/assets/pdf/GridMindDemo_JD4.mp4.

[31] RTE France and cha learn, learning to run a power network challenge[EB/OL]. [2020-04-29]. https://competitions.codalab.org/competitions/20767.

作者简介:

王之伟

王之伟(1966),男,硕士,高级工程师,IET会士,从事电力系统规划、智能调度、电力人工智能相关工作(E-mail: zhiwei.wang@geirina.net);

陆晓(1968),男,硕士,高级工程师,从事电力系统智能调度控制相关工作;

刁瑞盛(1981),男,博士,高级工程师,从事人工智能在电力系统中应用,智能调度系统研发相关工作。

Deep-reinforcement-learning based autonomous control and decision making for power systems

WANG Zhiwei[1,2], LU Xiao[2], DIAO Ruisheng[1], LI Haifeng[2], XU Chunlei[2], DUAN Jiajun[1], ZHANG Ningyu[3], SHI Di[1]

(1. Global Energy Interconnection Research Institute North America, San Jose 95134, USA;
2. State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210024, China;
3. State Grid Jiangsu Electric Power Co., Ltd. Research Institute, Nanjing 211103, China)

Abstract: Modern power grids are facing grand operational challenges due to highly intermittent and uncertain renewable energies as well as new types of loads, etc. In recent years, the rapid development of artificial intelligence (AI) technology has brought up new solutions for optimal control problems with high dimension, high nonlinearity and high dynamics. Based on deep reinforcement learning (DRL), a novel autonomous control platform is presented, which can realize online learning and decision making for power system dispatch and control. The target of the proposed control platform is to transform massive real-time measurements directly into control decisions within sub-second. In order to fully demonstrate the feasibility of the "grid mind", autonomous voltage control and line flow control are taken as two examples to formulate the methodology of DRL-based power system dispatch and control problem. Finally, both deep-Q-network and deep deterministic policy gradient algorithms are applied to demonstrate the strong learning capability of DRL agents and their effectiveness through extensive simulation results.

Keywords: artificial intelligence; grid mind; system dispatch and control; deep reinforcement learning; sub-second control

(编辑 胡昊明)