第48卷 第20期 2024年10月25日
DOI：10.7500/AEPS20240416006
电力系统自动化
Automation of Electric Power Systems
Vol. 48 No. 20 Oct. 25，2024

基于贝叶斯深度强化学习的主动配电网电压控制

张 晓[1]，吴 志[1]，郑 舒[1,2]，顾 伟[1]，胡 博[3]，董吉超[3]
（1. 东南大学电气工程学院，江苏省南京市 210096；
2. 南瑞集团有限公司（国网电力科学研究院有限公司），江苏省南京市 211106；
3. 国网辽宁省电力有限公司大连供电公司，辽宁省大连市 116001）

摘要： 多分布式源荷接入导致配电网电压波动性增强。同时，上级主网电压的不确定性波动也会对配电网电压特性造成影响。为有效应对主配电网电压波动，文中提出一种基于数据驱动与模型求解结合的主动配电网多时间尺度电压控制框架。在慢时间尺度下，考虑主网电压波动，构建了上级主网非无穷大系统多馈线环境，对该环境下的电压控制问题进行了对抗性马尔可夫过程建模。在训练过程中，以投影梯度下降算法使主网电压发生扰动。利用贝叶斯深度Q网络算法感知上级主网电压波动，实现有载调压变压器挡位快速控制。在快时间尺度下，基于传统二阶锥优化方法控制光伏逆变器无功输出。算例结果表明，该方法可准确感知上级主网电压波动，在极短时间实现配电网无模型电压控制，保证各节点电压维持在安全范围内。

关键词： 主动配电网；电压控制；多时间尺度；对抗性马尔可夫过程；投影梯度下降；贝叶斯深度Q网络；深度强化学习

0 引言

伴随大规模分布式资源的广泛接入，电力系统正经历结构性改革，向低碳化、分布式、智能化与开放化的方向发展[1-2]。其中，分布式能源的广泛接入在促进电力系统低碳化转型的同时，降低了系统短路比（short circuit ratio，SCR），使得配电网抗干扰能力下降，其主网电压也在一定程度上受分布式资源随机性、间歇性波动和输电网故障等因素的影响而波动。馈线内部新能源和主网电压的随机性波动导致配电网内部更易出现电压越限与电压不平衡的问题[3]。为应对上述主配电网的不确定性，亟须提出一种兼具鲁棒性和快速性的主动配电网电压控制策略，保障配电网的安全稳定运行。

主网难以预测的随机性与馈线内部分布式资源波动特性日益加剧，传统模型驱动算法已难以满足控制的实时性和鲁棒性需求[4]。目前，有较多研究借助具备强大数据挖掘能力和运算时效性的深度强化学习（deep reinforcement learning，DRL）技术[5]应对配电网运行控制过程中的不确定性。文献[6]提出一种基于图卷积神经网络的深度Q学习方法，提取配电网拓扑特征以应对实际运行过程中线路拓扑的频繁变化。文献[7]提出一种基于多智能体深度确定性策略梯度算法的分散式电压控制方法以应对光伏和风电机组引发的电压波动。文献[8-9]利用离线训练与在线训练结合的两阶段DRL算法，提升智能体对实际配电网线路参数的适应性。文献[10]将DRL算法与模型驱动算法结合，在智能体动作的下层构建基于光伏出力不确定性场景的随机优化模型以实现环境的不确定性建模。文献[11]建立了云-边协同的DRL分布式电源就地自适应控制框架以应对光伏不确定性。文献[12]建立了基于贝叶斯深度确定性策略梯度算法的配电网能量管理模型，增强了调度策略对极端情况的适应性。但目前尚未有研究利用DRL算法挖掘主网不确定性对配电网各馈线电压及控制策略的影响。

高比例新能源渗透提升了输配电网控制环境的复杂性，目前，已有文献利用输配电网联合仿真及优化分析该不确定性环境并提供优化控制方案。文献[13]构建了高比例新能源渗透公共连接点电压波动模型，在暂态层面分析了不同新能源渗透率下的电压波动特性。文献[14-15]论证了高比例新能源渗透对主网电压安全和波动特性造成的影响，分析了分布式新能源出力与主网电压的相关性。但上述研究并未考虑主网电压波动对配电网运行状态的影

收稿日期：2024-04-16；修回日期：2024-05-16。
上网日期：2024-07-24。
国家电网公司科技项目（5400-202328548A-3-2-ZN）。

响，也未针对上级主网电压波动提出有效的感知与控制策略。文献[16-18]建立了输配电网协同优化模型，利用模型驱动算法获取多能源最优调度和控制策略，但上述研究对模型和量测的依赖度较高，随着网络规模和控制设备的增多，时效性将显著降低。

为充分考虑主网电压不确定性对配电网馈线内部电压的影响并借助DRL算法挖掘主网电压不确定性与控制策略的联系，提升控制鲁棒性与时效性，本文在配电网电压控制过程中考虑上级主网电压波动，提出了一种基于数据驱动与模型求解结合的多时间尺度电压控制方法。首先，构建了考虑上级主网电压波动的配电网多馈线多时间尺度电压控制模型。在慢时间尺度下，对该环境中的电压控制问题进行对抗性马尔可夫过程建模，训练过程中引入投影梯度下降(projected gradient descent，PGD)算法使上级主网电压发生波动以模拟不确定性，利用贝叶斯深度Q网络(Bayesian deep Q network，BDQN)算法训练并控制有载调压变压器(on-load tap changer，OLTC)动作。在快时间尺度下，利用传统二阶锥优化方法控制光伏逆变器输出。

1　上级主网电压波动下的多馈线多时间尺度电压控制模型

1.1　上级主网电压波动的多馈线环境

通常，可根据SCR指标将配电系统强度分为3类，记SCR指标为K_{SCR}，当$K_{\mathrm{SCR}}\leqslant 2$时，系统极弱；当$2<K_{\mathrm{SCR}}\leqslant 3$时，系统弱；当$K_{\mathrm{SCR}}>3$时，系统强。系统强度越高，其抵御输电网故障、间歇性新能源波动和随机性负荷的能力就越强[19]。在理想情况下，变电站高压侧上级主网可视为无穷大系统，即进行最优潮流或负荷潮流计算时，配电网上级主网电压的标幺值恒定为1。

随着分布式新能源在配电系统中大规模集成，系统SCR指标持续下降，正经历由强到弱的转变。此时，配电系统上级主网不可视为无穷大电源，在主网与馈线间应接入阻抗模拟电源内阻，且上级主网电压受多重因素影响，难以预测波动特性。根据文献[19]，上级主网为非无穷大系统的配电网正常运行状态如附录A图A1所示。该系统K_{SCR}的计算式为[20]：

$$K_{\mathrm{SCR}}=\frac{S_{\mathrm{ac}}}{P_{\mathrm{N}}}=\frac{U_{\mathrm{N}}^2}{P_{\mathrm{N}}}\frac{1}{|Z_{\mathrm{s}}|} \tag{1}$$

式中：S_{ac}为变电站出口母线短路容量；Z_{s}为主网等效阻抗；P_{N}为系统额定功率；U_{N}为变电站出口额定电压。

根据式(1)，若给定配电网系统SCR的估计值，即可反推出主网等效阻抗，代入后续潮流分析。

在多馈线系统中，配电网利用变电站中的OLTC进行调压，对配电网馈线根节点电压进行调整。根据文献[21]，对主网与配电网根节点间OLTC支路与变电站出线支路进行建模，支路结构如附录A图A2所示，各物理量约束如式(2)所示。

$$\begin{cases} P_{\mathrm{m}}-L_{\mathrm{m}}r_{\mathrm{m}}+P_{\mathrm{e}}=\sum\limits_{f\in\chi}P_{\mathrm{e},f} \\ Q_{\mathrm{m}}-L_{\mathrm{m}}x_{\mathrm{m}}+Q_{\mathrm{e}}=\sum\limits_{f\in\chi}Q_{\mathrm{e},f} \\ U_{\mathrm{z}}^2-U_m^2=2(r_{\mathrm{m}}P_{\mathrm{m}}+x_{\mathrm{m}}Q_{\mathrm{m}})-(r_{\mathrm{m}}^2+x_{\mathrm{m}}^2)L_{\mathrm{m}} \\ L_{\mathrm{m}}=\dfrac{P_{\mathrm{m}}^2+Q_{\mathrm{m}}^2}{U_{\mathrm{z}}^2} \\ U_{\mathrm{m}}^2=d_{\mathrm{oltc}}^2U_{\mathrm{e}}^2 \end{cases} \tag{2}$$

式中：χ为以OLTC低压侧为首端的馈线支路集合；P_{m}和Q_{m}分别为主网传输至OLTC低压侧的有功和无功功率；P_{e}和Q_{e}分别为注入OLTC低压侧的有功和无功功率；$P_{\mathrm{e},f}$和$Q_{\mathrm{e},f}$分别为OLTC低压侧传输至配电网各馈线的有功和无功功率；r_{m}和x_{m}分别为主网与OLTC低压侧之间的电阻和电抗；L_{m}为主网至OLTC低压侧电流的平方；U_{z}、U_{e}、U_{m}分别为上级主网、OLTC低压侧和OLTC高压侧的电压；d_{oltc}为OLTC变比标幺值。

各非线性约束二阶锥松弛及线性化方法如附录B式(B1)和式(B2)所示。

1.2　双时间尺度电压控制模型

将传统慢时间尺度机电设备与新型电力电子设备相结合，建立双时间尺度电压控制模型。在每个慢时间尺度开始前根据日前光伏和负荷预测值控制OLTC挡位，待OLTC挡位确定后控制快时间尺度光伏逆变器无功功率输出。每个慢时间尺度的目标函数为最小化电压偏差，如式(3)所示。

$$\min\sum_{t=1}^{T_{\mathrm{s}}}\sum_{i=1}^{N}(U_{i,t}-1)^2 \tag{3}$$

式中：$U_{i,t}$为t时段节点i的电压；N为系统节点数；T_{s}为一个慢时间尺度中包含的快时间尺度调度总时段数。

基于Distflow潮流方程，各馈线物理量约束为：

$$\begin{cases} P_{i,t}=P_{i,t}^{\mathrm{pv}}+\sum\limits_{j\in F(i)}P_{ij,t}-\sum\limits_{k\in T(i)}(P_{ki,t}-l_{ki,t}r_{ki}) & i\in N_{\mathrm{S}} \\ Q_{i,t}=Q_{i,t}^{\mathrm{pv}}+\sum\limits_{j\in F(i)}Q_{ij,t}-\sum\limits_{k\in T(i)}(Q_{ki,t}-l_{ki,t}x_{ki}) & i\in N_{\mathrm{S}} \end{cases} \tag{4}$$

$$U_{j,t}^2=U_{i,t}^2-2(P_{ij,t}r_{ij}+Q_{ij,t}x_{ij})+l_{ij,t}(r_{ij}^2+x_{ij}^2) \qquad ij\in B_{\mathrm{S}} \tag{5}$$

$$l_{ij,t}U_{i,t}^2=P_{ij,t}^2+Q_{ij,t}^2 \qquad ij\in B_S \tag{6}$$

$$U_{\min}^2\leqslant U_{i,t}^2\leqslant U_{\max}^2 \tag{7}$$

式中：$P_{i,t}$和$Q_{i,t}$分别为t时段注入节点i的有功和无功功率；$P_{ij,t}$和$Q_{ij,t}$分别为t时段支路ij的有功和无功功率；$F(i)$为以节点i为首端点的支路的末端节点集合；$T(i)$为以节点i为末端节点的支路的首端节点集合；$l_{ki,t}$为t时段支路ki电流幅值的平方；$P_{i,t}^{pv}$为t时段节点i的光伏有功出力预测值；$Q_{i,t}^{pv}$为t时段节点i的光伏逆变器的无功输出；r_{ki}和x_{ki}分别为支路ki的电阻和电抗；N_S为系统节点集合；B_S为系统支路集合。

实际控制过程中，慢时间尺度OLTC挡位控制约束如式(2)所示；快时间尺度光伏有功与无功出力约束[11]如式(8)所示。

$$\begin{cases}0\leqslant P_{n,t}^{pv}\leqslant P_{n,t}^{pv,\max} & n\in N_{pv}\\ Q_{n,t}^{b}=\sqrt{m(S_n^{pv})^2-(P_{n,t}^{pv})^2} & n\in N_{pv}\\ -Q_{n,t}^{b}\leqslant Q_{n,t}^{pv}\leqslant Q_{n,t}^{b} & n\in N_{pv}\end{cases} \tag{8}$$

式中：N_{pv}为光伏机组接入点集合；n为光伏机组编号；$P_{n,t}^{pv}$为t时段光伏机组n的有功出力；$P_{n,t}^{pv,\max}$为t时段光伏机组n的有功出力上限；$Q_{n,t}^{pv}$为t时段光伏机组n的无功出力；$Q_{n,t}^{b}$为t时段光伏机组n的无功出力边界值；S_n^{pv}为光伏机组n的装机容量；m为容量弹性系数，即新能源无功出力可比实际值略高，本文取1.1。

2　数据驱动与模型求解结合的配电网电压控制框架

如图1所示，本文利用数据驱动算法控制慢时间尺度OLTC，利用模型求解算法控制快时间尺度光伏逆变器。在多馈线系统中，各馈线由于负荷和新能源渗透率不同，导致电压波动情况有差异。因此，在离线训练阶段，慢时间尺度OLTC智能体通过感知上级主网电压、各馈线历史负荷和新能源出力情况，基于对抗性马尔可夫过程进行经验积累和集中训练。具体而言，DRL智能体以30 min为时间尺度控制OLTC挡位，并将该动作与配电网环境交互以获取该步奖励值，奖励值为当前慢时间尺度动作下所有快时间尺度的最小电压偏差之和，求解采用基于模型求解的二阶锥规划(second-order cone programming，SOCP)方法以5 min为时间尺度控制光伏逆变器无功输出并获取最小电压偏差。当在线运行阶段的上级主网电压大幅波动时，训练完成的DRL智能体可在极短时间内配合各馈线具备实时响应功能的光伏逆变器，实现快速电压控制。

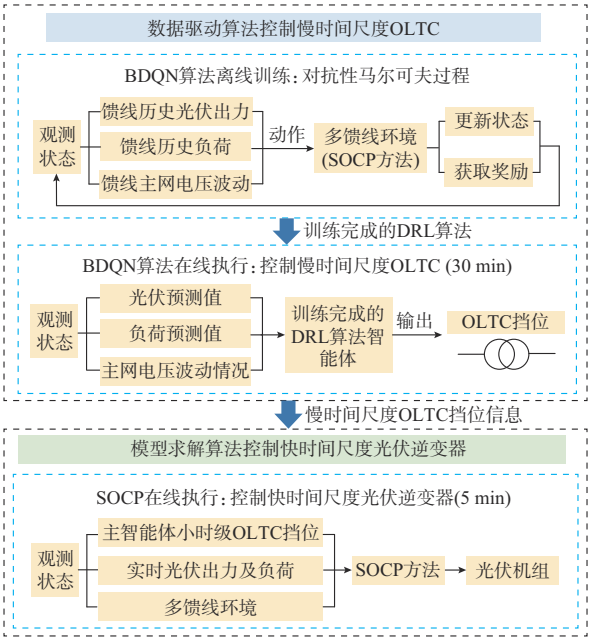

图1　多智能体多时间尺度控制框架
Fig. 1　Framework of multi-agent multi-timescale control

3　考虑上级主网波动的慢时间尺度DRL控制方法

3.1　慢时间尺度对抗性马尔可夫过程建模

3.1.1　对抗性马尔可夫过程

对抗性马尔可夫过程涉及智能体和对手两个主体，本质是一种马尔可夫博弈过程。对抗性马尔可夫过程可用一个5元数组(S,A,C,P,R)表示，相较于传统马尔可夫过程增加了对手扰动集C。其中，S为观测状态空间，A为动作集合，C为对手的状态扰动集合，P为状态转移概率，R为与环境交互获取的奖励集合。

在智能体固定策略函数π下，对手扰动函数$v(\cdot)$的目标为寻求最优扰动函数v^*，使得在当前策略函数π下的累计期望回报最小化，最优对手扰动下的Q函数$Q(\cdot)$为：

$$Q(s,a)=\min E\left[r+\gamma\max Q(v(s'),a')\right] \tag{9}$$

式中：γ为折扣系数；s和s'分别为当前和下一时刻状态；a和a'分别为智能体当前和下一时刻选择的动作；r为在对手扰动状态下，执行动作的奖励值；$E(\cdot)$为期望函数。

在最优扰动函数v^*下，智能体目标为寻找最优策略π^*，使得Q函数最大化，最优策略为：

$$\pi^*=\arg\max Q(s,a) \tag{10}$$

3.1.2　对抗性马尔可夫过程建模

在本文构建的上级主网电压波动的多馈线环境

下，可通过对慢时间尺度电压控制问题进行对抗性马尔可夫过程建模，在DRL训练过程中引入主网电压随机波动的干扰样本，以增强智能体应对实际控制过程中主网电压随机波动的能力。配电网慢时间尺度对抗性马尔过程如附录A图A3所示，智能体根据受对手扰动后的观测值输出OLTC动作并将其与弱电网环境交互，获取奖励值并更新观测状态。该对抗性马尔可夫过程的建模如下。

1)观测状态

智能体观测状态包括慢时间尺度内上级主网电压和各馈线各节点的净有功功率平均值。因此，在慢时间尺度内，智能体观测状态向量记为：$\boldsymbol{s}=\{U_{z,T},P_{1,T},P_{2,T},\cdots,P_{i,T},\cdots,P_{N,T}\}$，其中，$U_{z,T}$为$T$时段上级主网的电压幅值，$P_{i,T}$为$T$时段节点$i$的有功功率平均值。

2)对手扰动

对手扰动c为浮点数型连续变量，取值区间为$[-\Delta U,\Delta U]$，其中，ΔU为主网电压最大波动值。对手对智能体的观测状态进行扰动，使其观测到的上级主网电压在区间$[U_z-\Delta U,U_z+\Delta U]$内波动。训练过程中，以式(9)为目标，基于PGD算法获取最优扰动变量的值。

3)动作

OLTC动作a为整数型离散变量，本文设OLTC共计33个挡位[22-23]。因此，变量取值区间为$[-16,16]$，每个挡位可调节0.003 125 p.u.，合计共0.1 p.u.。

4)奖励值

慢时间尺度的单步奖励值定义为一个慢时间尺度调度周期内所有快时间尺度的节点电压偏差之和r_T：

$$r_T=-\sum_{t=1}^{T_s}\sum_{i=1}^{N}(U_{i,t}-1)^2 \tag{11}$$

快时间尺度电压偏差计算基于约束式(4)—式(8)，目标函数为最小化电压偏差，利用传统SOCP模型求解。

3.2　基于PGD算法的上级主网最优扰动量选取方法

本文构建的慢时间尺度对抗性马尔可夫过程中，对手利用最强一阶扰动——PGD算法求取对智能体上级主网电压观测量的最优扰动。

给定样本$(\boldsymbol{s},\boldsymbol{q})$和需要拟合的策略网络$f_\pi$，其中，$\boldsymbol{q}$为所有动作对应的Q值向量，对手的目标为在$\boldsymbol{s}$增加扰动后，使得新Q值向量与$\boldsymbol{q}$之间的交叉熵损失最低，记为：

$$\max_{\boldsymbol{c}} L(f_\pi(\boldsymbol{s}+\boldsymbol{c}),\boldsymbol{q}) \tag{12}$$

$$\boldsymbol{c}=[c,0,0,\cdots,0] \tag{13}$$

式中：$L(\cdot)$为交叉熵损失函数；c为对手扰动变量增广后的$N+1$维行向量。

PGD算法可协助对手寻求最优扰动，通过小步梯度下降和投影映射对$\boldsymbol{s}$进行干扰。将该算法应用到慢时间尺度智能体的训练过程中，每一训练步通过n次迭代攻击调整上级主网电压波动值，使价值网络输出的Q值分布与原始未受干扰的Q值分布相似度最低，即交叉熵损失函数最小。算法原理及伪代码见文献[24]与附录A。

3.3　贝叶斯DRL算法

基于3.1节和3.2节的描述，为应对实际控制过程中上级主网电压的不确定性，本文在慢时间尺度训练过程中利用PGD算法引入大量干扰样本，增大了智能体状态空间与学习难度，使得传统确定性DRL算法难以准确刻画波动状态与动作价值分布间的关系。本节引入不确定感知与泛化能力很强的贝叶斯DRL算法[25]对慢时间尺度电压控制问题进行训练，首先，介绍贝叶斯神经网络(Bayesian neural network，BNN)算法。

3.3.1　BNN算法

在贝叶斯DRL算法中，通常采用贝叶斯概率网络代替传统确定性神经网络以获取智能体观测状态对应的动作Q值分布[26]。与传统神经网络各神经元之间权重为定值不同，BNN算法各神经元之间的权重为符合正态分布的随机数，训练参数为各正态分布的均值与方差，具有不确定性量化能力。因此，BNN算法泛化能力更强，可提升智能体应对不确定性的能力，对本文所构建的考虑上级主网电压随机波动及分布式资源出力不确定性的配电网有更强的适应性。利用贝叶斯DRL算法对传统深度Q网络(deep Q network，DQN)算法进行改进，将深度确定性Q网络替换为BNN算法，可实现动作价值函数的不确定性估计。本文BNN算法的输入为慢时间尺度智能体受扰动后的观测状态，输出为33种挡位对应Q值的分布。

3.3.2　贝叶斯DRL算法原理

贝叶斯学习的训练目的为近似后验分布，在DRL学习过程中，其目标为给出准确的神经网络参数以近似正确的动作Q值分布[27]，预测分布的计算式为：

$$P_1(w|M)=\frac{P_2(w)P_3(M|w)}{P_4(M)} \tag{14}$$

式中：w为神经网络参数；M为输入与输出Q值的数据集合；$P_2(w)$为网络参数的先验概率；$P_1(w|M)$为网络参数准确的后验概率；$P_3(M|w)$为似然函数值；$P_4(M)$为边缘似然函数值。

在此基础上，利用变分推理和蒙特卡洛采样方法推导出贝叶斯神经网络损失函数，具体推导过程见附录B式(B3)—式(B7)，其中，式(B7)为最终结果。

附录B式(B7)中，第1项与第3项可通过对BNN算法各层级权重进行对数运算求得，第2项为计算数据集似然函数值，需要根据DRL算法的动作Q值分布获取原理进行推导，本文在3.3.3节中对BDQN算法的损失函数进行详细的推导。

3.3.3　BDQN算法

与文献[28]中所述DQN算法相同，BDQN算法采用3种机制增强学习性能，分别为经验回放、ε-贪婪策略和目标网络机制。在传统DQN算法损失函数的基础上，根据3.3.2节贝叶斯DRL算法原理推导出贝叶斯Q评估网络的损失函数。

1)经验回放机制

设置经验回放池D，由于受到对手攻击，当前时间步的状态回放与智能体所选动作受对手攻击影响，存放经验为$(\boldsymbol{v}(\boldsymbol{s}), a, r, \boldsymbol{s}')$，其中，$\boldsymbol{v}(\boldsymbol{s})$为经对手扰动后的状态向量，$\boldsymbol{s}'$为下一时刻状态向量。

2)ε-贪婪策略

为确保智能体能够充分探索，训练过程中引入ε-贪婪策略选择动作，即智能体探索幅度随动作次数增加而不断降低，具体公式参照文献[28]。

3)目标Q网络机制

训练时，Q评估网络的参数向量为$\boldsymbol{W}$，引入目标Q网络，其网络结构和参数向量$\boldsymbol{W}'$初始值与Q评估网络完全一致。在训练过程中，Q评估网络参数随智能体学习进行实时更新，在经过一定次数的更新后，评估网络将参数赋值给目标网络。传统DQN算法的Q评估网络损失函数推导过程见文献[28]，本文在此基础上推导出BDQN算法Q评估网络参数单步损失函数$H(\boldsymbol{W})$，如式(15)所示，具体推导过程见附录B式(B8)—式(B10)。

$$H(\boldsymbol{W})=\frac{1}{S}\sum_{i'=1}^{S}\sum_{n'=1}^{2N_p}\ln P_1(w_{n',i'}|W_{n',i'})-\frac{1}{S}\sum_{i'=1}^{S}\left\{-\frac{[r_{i'}+\gamma Q_{n,i'}(\boldsymbol{v}(\boldsymbol{s}'),a';\boldsymbol{W}')-Q_{n,i'}(\boldsymbol{v}(\boldsymbol{s}),a;\boldsymbol{W})]^2}{2\kappa^2}-\ln\kappa-\ln\sqrt{2\pi}\right\}-\frac{1}{S}\sum_{i'=1}^{S}\sum_{n'=1}^{2N_p}\ln P_2(w_{n',i'}) \tag{15}$$

式中：S为蒙特卡洛采样次数；$2N_p$为BNN算法参数的总数；$r_{i'}$为第i'次蒙特卡洛采样的奖励值；$Q_{n,i'}(\boldsymbol{v}(\boldsymbol{s}'),a';\boldsymbol{W}')$为参数为$\boldsymbol{W}'$的Q评估网络第$i'$次蒙特卡洛采样时输入下一时刻状态向量$\boldsymbol{v}(\boldsymbol{s}')$并选择动作$a'$情况下的Q值；$Q_{n,i'}(\boldsymbol{v}(\boldsymbol{s}),a;\boldsymbol{W})$为参数为$\boldsymbol{W}$的Q评估网络第$i'$次蒙特卡洛采样时输入当前时刻状态向量$\boldsymbol{v}(\boldsymbol{s})$并选择动作$a$情况下的Q值；$\kappa$为正态分布方差噪声变量，用于增强似然计算的鲁棒性；$P_1(w_{n',i'}|W_{n',i'})$为Q评估网络第i'次蒙特卡洛采样第n'个参数的后验概率值；$P_2(w_{n',i'})$为Q评估网络第i'次蒙特卡洛采样第n'个参数的先验概率值。

综上所述，慢时间尺度BDQN算法流程如附录A所示。训练完成后的BDQN算法将被用于控制慢时间尺度OLTC挡位，智能体根据上级主网实际电压波动和平均功率的观测输出33个挡位对应的Q值，然后选取Q值最大的挡位作为智能体输出动作。

4　算例分析

4.1　算例介绍

本文仿真测试基于硬件平台NVIDIA GTX 4060Ti和Intel Core i5-13490KF实现；软件系统为Windows 10；DRL算法均基于Python 3.9.0的Pytorch神经网络框架训练实现。

选取某地区实际100节点三馈线配电网系统进行分析，如附录A图A4所示。上级主网电压为110 kV，经过变电站变压为10 kV，额定视在功率为100 MV·A。在本文所构建的配电网模型中，变压器两端均有参数，采用文献[29]所述第2种标幺化方式，将线路及各设备参数的有名值除以由基本级归算到这些物理量所在电压等级的基准值以进行就地标幺化。各光伏机组安装位置及容量如附录A表A1所示，光伏总装机容量为19.35 MW。训练集与测试集数据见文献[30]。

为保证公平性，在进行方法对比时，算例分析时选用的DQN算法超参数及网络结构等设定与本文提出的BDQN算法保持一致，PGD算法参数、BDQN算法参数及训练超参数设定如附录A表A2所示。其中，PGD算法最大扰动参数设定为0.05，即上级主网电压波动区间为[−0.05, 0.05]p.u.。该多馈线系统各节点电压安全运行范围为[0.95, 1.05]。

4.2　测试集电压偏差情况分析

首先，对测试集配电网环境电压偏差情况进行分析。当上级主网电压在[−0.05, 0.05]p.u.范围内随机波动时，不考虑任何电压控制设备，基于最优潮流技术，分析无电压控制设备时该系统电压偏差问题。以5 min为一个时间断面进行分析，测试集共计288个时间断面，各节点288个时间点电压标幺值分别如附录A图A5和图A6所示。节点50至70和节点75至79在时段250至280出现了严重的电压跌落现象，测试集全天电压最小值为0.911 p.u.，原因在于该时段严重的源荷不匹配现象，即负荷需求量大而光伏有功出力不足，亟须传统控制设备与新型电力电子设备共同作用以抬升电压。

4.3　贝叶斯DRL算法效果分析

为验证本文所提方法在范围较大的多馈线系统中应对上级主网电压波动的有效性，将本文所提方法（方法1）与两种DRL算法进行对比，具体如表1所示。

表1　不同方法的设置
Table 1　Settings for different methods

方法编号	马尔可夫过程	训练时扰动情况	训练算法	训练参数
方法1(本文所提)	对抗性	PGD	BDQN	保持一致
方法2	对抗性	PGD	DQN	
方法3	普通	无扰动	BDQN	

4.3.1　算法收敛性分析

基于相同的超参数训练方法1和方法3智能体。每回合包括48个智能体动作周期和288个光伏逆变器动作周期，回合训练开始前，随机抽取训练集中的一天进行训练，训练400回合后结束。两种算法的收敛情况如图2所示。图中的蓝色曲线表示方法1的训练情况。在训练初期，由于智能体探索和PGD算法的影响，奖励值较低且大幅波动，但总体呈现上升趋势。随着回合数的增加，经验回放池填满，智能体探索幅度下降，BDQN算法依据大量的经验回放感知上级主网电压不确定性，奖励值不断上升，在大约50回合后收敛。收敛后的奖励值呈小幅上升趋势，最终收敛到−3左右，算法收敛稳定，具备应对扰动的能力。图中红色曲线表示的DQN算法虽然开始时奖励值高于BDQN算法，但由于网络缺乏应对不确定性的感知与泛化能力，奖励值虽较快达到−3，但在PGD算法的作用下频繁出现较大扰动，选择次优解甚至较差解，难以适应上级主网电压波动。

4.3.2　算法电压控制有效性分析

为验证所提算法在主网电压随机波动下的有效性，本节对方法1中训练完成的BDQN算法智能体电压控制效果进行测试。首先，分析算法在主网电压分别以0%、10%、20%、30%、40%和50%概率的PGD算法扰动下的电压控制效果，测试集全天电压偏差值、最大值和最小值如附录A表A3所示。需要注意的是，BDQN算法的Q网络为BNN算法，网络参数为服从特定正态分布的随机数，故对应一固定观测状态，每次前向传播输出的动作Q值分布可能有微小差异。因此，在实际决策时，本文选取10次采样的平均值作为最终动作。

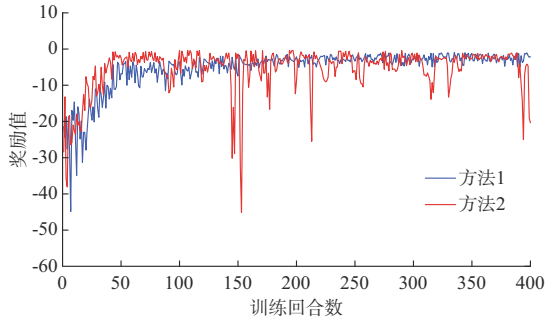

图2　奖励值的收敛曲线
Fig. 2　Convergence curves of reward values

由附录A表A3可知，当主网电压以不同概率在区间[−0.05, 0.05]p.u.波动时，方法1可将各节点电压维持在安全范围内。随着波动概率的上升，测试集全天电压偏差稳定，无显著上升趋势，表明BDQN算法智能体能准确感知上级主网电压波动，控制OLTC和光伏逆变器合理动作。

验证上级主网电压以90%概率受PGD算法扰动的极端情况下方法1的有效性，主网电压全天波动情况和OLTC挡位动作信息如图3所示，测试集全天电压控制情况如附录A图A7所示。

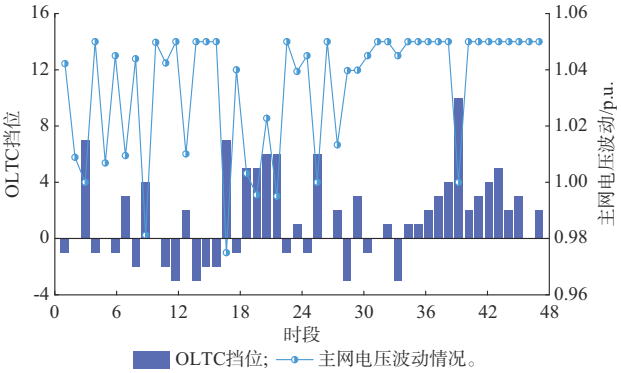

图3　主网电压极端波动情况和OLTC动作挡位
Fig. 3　Extreme fluctuations of main grid voltage and OLTC tap

经计算，测试集全天电压偏差为3.281 p.u.，电压最大值为1.037 p.u.，最小值为0.961 p.u.，均处于

安全范围内。由图3和图A7可知，在时段0至3、5至8、12至16、32至39和41至48时，上级主网电压上偏严重，故OLTC智能体挡位设置较低，防止多馈线各节点出现电压越界现象；在4、9、17和40时段，上级主网电压下偏严重，OLTC挡位随之抬升，以缓解电压跌落。

4.3.3　训练期间引入PGD算法效果分析

为验证在训练过程中基于对抗性马尔可夫过程对智能体观测状态引入PGD算法对主网电压进行扰动的必要性，对方法1和方法3的算法收敛情况和电压控制效果进行分析。两者的训练奖励值收敛曲线见附录A图A8，测试过程中，主网电压受不同概率的PGD算法扰动而波动，方法3的BDQN算法全天电压偏差结果如附录A图A9所示。

综合分析附录A图A8、图A9及表A3可知，BDQN算法具有极强的泛化能力，无论训练过程中是否增加对上级主网电压的扰动，其奖励曲线收敛性较好且相近。方法3的BDQN算法测试集全天电压偏差值显著低于方法1，且随着主网电压波动概率的上升，测试集全天电压偏差呈显著上升趋势。当扰动概率不小于40%时，节点电压最小值出现越界现象，最大值有显著的越界趋势，不能充分感知主网电压波动。因此，在应对难以预测的上级主网电压不确定性时，基于对抗性马尔可夫过程训练智能体可显著提升算法应对扰动的鲁棒性。

4.4　不同方法控制效果对比

为验证本文所提基于对抗性马尔可夫过程训练的BDQN算法的实际控制优势，对比表1中3种方法和传统多时间尺度模型驱动方法在上级主网电压以90%概率波动下的电压控制效果，包括测试集全天平均电压偏差、电压最大值、电压最小值以及慢时间尺度OLTC动作平均决策时长，如表2所示。其中传统多时间尺度模型控制方法的上级主网电压波动情况与方法1保持一致，目标函数为式(3)，约束条件为式(2)、式(4)—式(8)，利用Gurobi求解器求解。

表2　不同方法的电压控制效果对比
Table 2　Comparison of voltage control effects of different methods

算法	平均偏差/p.u.	最大值/p.u.	最小值/p.u.	训练时长/min	执行时长/s
方法1	0.011 3	1.037 5	0.960 4	192	0.002 4
方法2	0.024 1	1.045 3	0.968 9	185	0.002 2
方法3	0.022 5	1.033 4	0.936 6	285	0.001 8
传统	0.009 7	1.031 3	0.962 8		7.263 0
无控制	0.034 0	1.022 5	0.924 5		

由表2可知，方法1的测试集平均电压偏差与传统求解器所得全局最优电压偏差接近，按式(16)计算[11]，其电压控制最优率为93.41%，逼近全局最优。

$$\eta=\left(1-\left|\frac{\delta_{\mathrm{p}}-\delta_{\mathrm{c}}}{\delta_{\mathrm{c}}-\delta_{\mathrm{o}}}\right|\right)\times 100\% \tag{16}$$

式中：δ_{p}、δ_{c}和δ_{o}分别为本文所提控制、传统最优控制和无电压控制3种情况下的平均电压偏差。

方法1的全天电压最大值为1.037 5 p.u.，最小值为0.960 4 p.u.，能够在上级主网电压大范围波动情况下维持测试集全天各节点电压在安全范围内，且其电压偏差显著低于方法2与方法3，证明其相较于非对抗性马尔可夫过程下的DRL算法和DQN算法有更强的鲁棒性。

更进一步，选取部分光伏接入节点及支路末端节点对比方法1与传统求解算法的电压控制效果，节点16和节点69的电压控制效果如图4所示，节点33和节点90的电压控制效果如附录A图A10所示，两种方法各关键节点电压控制效果接近。图4中，方法1计算节点69的电压在第200个时段附近与传统算法求解结果相差0.01 p.u.，原因为DRL算法通过智能体与环境的交互训练更新神经网络参数以逼近全局最低电压偏差，但很难保证每个节点在每个时段的电压控制效果均与传统控制一致。经计算，图4所示方法1控制的节点69全天电压总偏差为0.065 p.u.，相较于传统求解算法的0.069 p.u.降低了0.004 p.u.，说明本文所提DRL算法可更有效地控制支路末端节点电压。

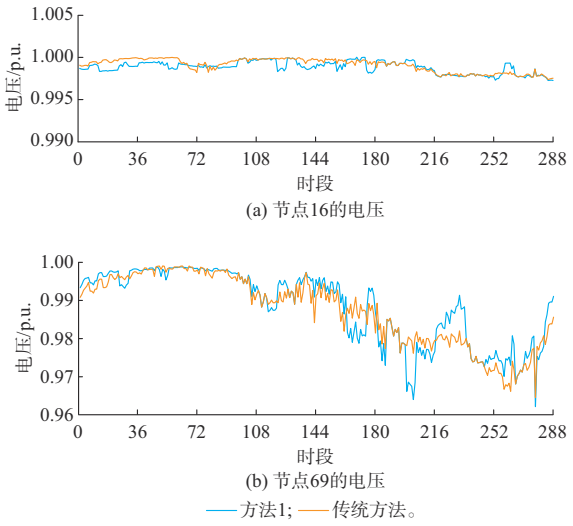

图4　节点电压对比
Fig. 4　Comparison of node voltage

除了近似最优的控制效果，本文所提方法中一次慢时间尺度OLTC平均决策计算时间（即4.3.2节中所述利用训练完成的DRL算法“采样10次并取平均值”所耗时间）为0.002 4 s。一次慢时间尺度全部设备的平均优化决策计算时间，包括OLTC挡位的DRL决策和光伏逆变器的SOCP方法优化时间，为0.608 s，相较于传统求解算法的7.263 s，在时效性方面有显著提升，在面临极端波动情况时，能以极快的速度响应上级主网电压波动，避免各节点电压出现越界可能，保障配电网安全稳定运行。

5　结语

本文针对多分布式资源接入背景下的主动配电网提出一种考虑上级主网电压波动的实时电压控制方法，实现了主配电网不确定性环境下多馈线多时间尺度电压控制。主要结论如下：

1）构建了上级主网为非无穷大系统及主网电压不确定性波动的多馈线配电网电压控制环境，相较于理想模型考虑了输电网不确定性，更符合多分布式资源接入的配电网实际运行环境。

2）构建了数据驱动与模型求解结合的配电网多智能体多时间尺度电压控制框架，利用贝叶斯DRL算法控制慢时间尺度OLTC挡位，利用传统SOCP方法控制快时间尺度光伏逆变器的无功输出。

3）对上级主网电压波动下的配电网慢时间尺度电压控制问题进行对抗性马尔可夫过程建模，并利用BDQN算法进行训练与控制；在训练过程中引入PGD算法使上级主网电压发生扰动，以提升智能体实际控制过程中对主网电压不确定性波动的感知的应对能力。

4）实际100节点三馈线系统测试结果表明，在控制效果方面，相较于DQN算法和训练过程中未加入扰动的BDQN算法，本文所提方法在上级主网电压以90%概率在[−0.05, 0.05]p.u.波动的极端情况下可实现有效无模型电压控制，平均电压偏差、电压最大值和最小值与传统最优控制结果接近，控制最优率为93.41%；在决策时间方面，本文所提方法慢时间尺度OLTC挡位平均决策时长为0.002 4 s，相较于传统方法的7.263 s，时效性有显著提升。

本文的模型尚未考虑配电网拓扑的不确定性，未来将研究配电网多馈线拓扑重构下的不确定性特征提取及拓扑重构与多种设备的协同电压控制方法。

附录见本刊网络版（http://www.aeps-info.com/aeps/ch/index.aspx），扫英文摘要后二维码可以阅读网络全文。

参考文献

[1] 魏泓屹，卓振宇，张宁，等. 中国电力系统碳达峰·碳中和转型路径优化与影响因素分析[J]. 电力系统自动化，2022，46(19)：1-12.
WEI Hongyi, ZHUO Zhenyu, ZHANG Ning, et al. Transition path optimization and influencing factor analysis of carbon emission peak and carbon neutrality for power system of China [J]. Automation of Electric Power Systems, 2022, 46(19): 1-12.

[2] 李鹏，习伟，李鹏，等. 基于边缘计算的配电网数字化转型关键问题分析与展望[J]. 电力系统自动化，2024，48(6)：29-41.
LI Peng, XI Wei, LI Peng, et al. Analysis of key issues and prospect for digital transformation of distribution networks based on edge computing[J]. Automation of Electric Power Systems, 2024, 48(6): 29-41.

[3] 杨浩，苏文栋，谷毅，等. 面向耦合系统的交替方向滚动时域电压分层协同优化控制[J]. 电工技术学报，2023，38(21)：5846-5861.
YANG Hao, SU Wendong, GU Yi, et al. Voltage hierarchical cooperative control of coupled system using model predictive control and alternating direction method of multipliers [J]. Transactions of China Electrotechnical Society, 2023, 38(21): 5846-5861.

[4] WU Z, LI Y Q, GU W, et al. Multi-timescale voltage control for distribution system based on multi-agent deep reinforcement learning[J]. International Journal of Electrical Power & Energy Systems, 2023, 147: 108830.

[5] 胡维昊，曹迪，黄琦，等. 深度强化学习在配电网优化运行中的应用[J]. 电力系统自动化，2023，47(14)：174-191.
HU Weihao, CAO Di, HUANG Qi, et al. Application of deep reinforcement learning in optimal operation of distribution network [J]. Automation of Electric Power Systems, 2023, 47(14): 174-191.

[6] HOSSAIN R R, HUANG Q H, HUANG R K. Graph convolutional network-based topology embedded deep reinforcement learning for voltage stability control [J]. IEEE Transactions on Power Systems, 2021, 36(5): 4848-4851.

[7] 徐博涵，向月，潘力，等. 基于深度强化学习的含高比例可再生能源配电网就地分散式电压管控方法[J]. 电力系统保护与控制，2022，50(22)：100-109.
XU Bohan, XIANG Yue, PAN Li, et al. Local decentralized voltage management of a distribution network with a high proportion of renewable energy based on deep reinforcement learning [J]. Power System Protection and Control, 2022, 50(22): 100-109.

[8] PETRUSEV A, PUTRATAMA M A, RIGO-MARIANI R, et al. Reinforcement learning for robust voltage control in distribution grids under uncertainties [J]. Sustainable Energy, Grids and Networks, 2023, 33: 100959.

[9] LIU H T, WU W C. Two-stage deep reinforcement learning for inverter-based volt-var control in active distribution networks[J]. IEEE Transactions on Smart Grid, 2021, 12(3): 2037-2047.

[10] LIU H Y, ZHANG C, CHAI Q M, et al. Robust regional coordination of inverter-based volt/var control via multi-agent

deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2021, 12(6): 5420-5433.

[11] 习伟，李鹏，李鹏，等.基于深度强化学习的分布式电源就地自适应电压控制方法[J].电力系统自动化，2022，46(22)：25-31.
XI Wei, LI Peng, LI Peng, et al. Adaptive local voltage control method for distributed generator based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2022, 46(22): 25-31.

[12] ZHANG T Q, SUN M Y, QIU D W, et al. A Bayesian deep reinforcement learning-based resilient control for multi-energy micro-gird[J]. IEEE Transactions on Power Systems, 2023, 38(6): 5057-5072.

[13] BRINKEL N B G, GERRITSMA M K, ALSKAIF T A, et al. Impact of rapid PV fluctuations on power quality in the low-voltage grid and mitigation strategies using electric vehicles[J]. International Journal of Electrical Power & Energy Systems, 2020, 118: 105741.

[14] ZHAO W L, SHI X C, GUO Q L, et al. Voltage security analysis with high PVs penetration considering the interaction of transmission and distribution grids: case studies[C]// 2015 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), November 15-18, 2015, Brisbane, Australia.

[15] CHEN Z H, YU R J, HU C G, et al. The voltage fluctuation characteristic analysis of regional power network with high-penetration PV system [C]// 2017 4th International Conference on Systems and Informatics (ICSAI), November 11-13, 2017, Hangzhou, China.

[16] 路镇铭，徐英，仪忠凯，等.基于双向安德森加速分解协调算法的输配协同无功优化[J/OL].电力系统自动化[2024-01-29]. https://kns.cnki.net/kcms/detail/32.1180.TP.20240126.2057.003.html.
LU Zhenming, XU Ying, YI Zhongkai, et al. Transmission and distribution coordinated reactive power optimization based on bidirectional Anderson acceleration decomposition coordination algorithm[J/OL]. Automation of Electric Power Systems [2024-01-29]. https://kns. cnki. net/kcms/detail/32.1180.TP.20240126.2057.003.html.

[17] 刘晨曦，辛焕海.考虑实际工况的新能源多馈入系统广义运行短路比[J/OL].电力系统自动化[2024-04-18].https://kns.cnki.net/kcms/detail/32.1180.tp.20240415.2009.010.html.
LIU Chenxi, XIN Huanhai. Generalized operational short-circuit ratio of multi-infeed systems with renewable energy considering actual operation conditions[J/OL]. Automation of Electric Power Systems [2024-04-18]. https://kns. cnki. net/kcms/detail/32.1180.tp.20240415.2009.010.html.

[18] 华文，董炜，阙凌燕，等.计及直流互联与电-气-热耦合的输配协同优化调度[J].浙江电力，2022，41(11)：31-38.
HUA Wen, DONG Wei, QUE Lingyan, et al. Collaborative optimal scheduling for coupled transmission and distribution systems considering HVDC interconnection and electricity-gas-heat coupling[J]. Zhejiang Electric Power, 2022, 41(11): 31-38.

[19] 李胜男，高源，何廷一，等.新能源临界短路比对电网运行参数灵敏度分析方法[J].智慧电力，2023，51(5)：96-102.
LI Shengnan, GAO Yuan, HE Tingyi, et al. Sensitivity analysis method for critical short circuit ratio of new energy to power grid operating parameters[J]. Smart Power, 2023, 51(5): 96-102.

[20] 王奕鑫，唐宇，叶琳，等.多馈入直流系统电压稳定中静态负荷影响的定量分析和评估方法[J/OL].电工技术学报[2024-05-15].https://doi.org/10.19595/j.cnki.1000-6753.tces.231870.
WANG Yixin, TANG Yu, YE Lin, et al. Quantitative analysis and evaluation method for the impact of static loads on voltage stability in multi-infeed direct current systems[J/OL]. Transactions of China Electrotechnical Society [2024-05-15]. https://doi.org/10.19595/j.cnki.1000-6753.tces.231870.

[21] 吴文传，张伯明，巨云涛.主动配电网网络分析与运行调控[M].北京：科学出版社，2016.
WU Wenchuan, ZHANG Boming, JU Yuntao. Active distribution networks analysis, operation and control [M]. Beijing: Science Press, 2016.

[22] WANG R H, BI X W, BU S Q. Real-time coordination of dynamic network reconfiguration and volt-var control in active distribution network: a graph-aware deep reinforcement learning approach[J]. IEEE Transactions on Smart Grid, 2024, 15(3): 3288-3302.

[23] XU H C, DOMINGUEZ-GARCIA A D, SAUER P W. Optimal tap setting of voltage regulation transformers using batch reinforcement learning[J]. IEEE Transactions on Power Systems, 2020, 35(3): 1990-2001.

[24] WANG C. Attacking and defending deep reinforcement learning policies [EB/OL]. [2024-01-29]. https://arxiv. org/abs/2205.07626.

[25] 冯昌森，张瑜，文福拴，等.基于深度期望Q网络算法的微电网能量管理策略[J].电力系统自动化，2022，46(3)：14-22.
FENG Changsen, ZHANG Yu, WEN Fushuan, et al. Energy management strategy for microgrid based on deep expected Q network algorithm[J]. Automation of Electric Power Systems, 2022, 46(3): 14-22.

[26] GHAVAMZADEH M, MANNOR S, PINEAU J, et al. Bayesian reinforcement learning: a survey[EB/OL]. [2024-04-16]. http://arxiv.org/abs/1609.0443644.

[27] SILVA B D. A Bayesian perspective on Q-learning[EB/OL]. [2024-04-16]. https://brandinho. github. io/bayesian-perspective-q-learning.

[28] 李宏浩，张沛，刘墨.基于深度强化学习的暂态稳定紧急控制决策方法[J].电力系统自动化，2023，47(5)：144-152.
LI Honghao, ZHANG Pei, LIU Zhao. Decision-making method for transient stability emergency control based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2023, 47(5): 144-152.

[29] 陈珩.电力系统稳态分析[M].北京：中国电力出版社，2007.
CHEN Heng. Steady state analysis of power systems [M]. Beijing: China Electric Power Press, 2007.

[30] BDQN training and testing set [EB/OL]. [2024-04-16]. https://github.com/chilixiao/BDQN-training-and-testing-set.

张　晓(2000—)，女，硕士研究生，主要研究方向：配电网电压控制。E-mail：1838509583@qq.com

吴　志(1987—),男,通信作者,博士,副教授,主要研究方向:输电网及配电网运行与规划、分布式发电、电力市场。E-mail:zwu@seu.edu.cn

郑　舒(1989—),女,硕士,工程师,主要研究方向:交直流混合配电网运行控制关键技术。E-mail:carmeng89@126.com

(编辑　杨松迎)

Voltage Control for Active Distribution Network Based on Bayesian Deep Reinforcement Learning

ZHANG Xiao[1], *WU Zhi*[1], *ZHENG Shu*[1,2], *GU Wei*[1], *HU Bo*[3], *DONG Jichao*[3]

(1. School of Electrical Engineering, Southeast University, Nanjing 210096, China;
2. NARI Group Corporation (State Grid Electric Power Research Institute), Nanjing 211106, China;
3. Dalian Electric Power Supply Company of State Grid Liaoning Electric Power Co., Ltd., Dalian 116001, China)

Abstract: The access of multiple distributed sources and loads leads to enhanced voltage volatility in the distribution network. Meanwhile, the uncertainty fluctuation in the voltage of the upper main grid also affect the voltage characteristics of the distribution network. In order to effectively deal with the voltage fluctuations of the main grid and the distribution network, this paper proposes a multi-timescale voltage control framework for active distribution networks based on the combination of data-driven and model solving. In the slow time scale, considering the voltage fluctuation of the main grid, a multiple-feeder environment with a non-infinity system of the upper main grid is constructed, and the voltage control problem in this environment is modeled as an adversarial Markov process. During the training process, the voltage of the main grid is perturbed with a projected gradient descent algorithm. The Bayesian deep Q network algorithm is utilized to sense the voltage fluctuation of the upper main grid and realize the fast control of taps of the on-load tap changer. In the fast time scale, the reactive power output of the photovoltaic inverter is controlled based on the traditional second-order cone optimization method. The case results show that the method can accurately sense the voltage fluctuation of the upper main grid, realize model-free voltage control of the distribution network in a very short time, and ensure that the voltage of each node is maintained within the safety range.

This work is supported by State Grid Corporation of China (No. 5400-202328548A-3-2-ZN).

Key words: active distribution network; voltage control; multi-timescale; adversarial Markov process; projected gradient descent; Bayesian deep Q network; deep reinforcement learning