电力系统自动化
Automation of Electric Power Systems
ISSN 1000-1026,CN 32-1180/TP

《电力系统自动化》网络首发论文

题目：配电系统双时间尺度电压管理的深度强化学习方法
作者：冯昌森，张瑜，谢路耀，文福拴，张凯怡，张有兵
收稿日期：2021-12-20
网络首发日期：2022-03-23
引用格式：冯昌森，张瑜，谢路耀，文福拴，张凯怡，张有兵．配电系统双时间尺度电压管理的深度强化学习方法[J/OL]．电力系统自动化.
https://kns.cnki.net/kcms/detail/32.1180.tp.20220321.1656.010.html

第XX卷 第XX期 XXXX年XX月XX日
DOI：10.7500/AEPS20211220001
电力系统自动化
Automation of Electric Power Systems
Vol. XX No. XX XX XX, XXXX

配电系统双时间尺度电压管理的深度强化学习方法

冯昌森[1]，张　瑜[1]，谢路耀[1]，文福拴[2]，张凯怡[1]，张有兵[1]
（1. 浙江工业大学信息工程学院，浙江省杭州市 310023；2. 浙江大学电气工程学院，浙江省杭州市 310027）

摘要：随着可再生能源发电渗透率的不断增大，配电系统的电压越限的问题愈发频繁，亟需高效的电压管理策略以保证配电系统的安全经济运行。首先，文中建立了双时间尺度的配电系统电压管理模型，实现不同时间响应特性的调压设备协调控制。然后，将2个时间尺度的电压管理模型建模为马尔可夫决策过程，在有效考虑两者的时间耦合关系和可控设备物理特性的基础上，分别利用多智能体深度确定性策略梯度算法和双深度Q网络算法求解模型实现了双时间尺度的实时电压管理。最后，基于IEEE 33节点配电系统进行算例分析，验证了所提模型和方法的有效性。

关键词：配电系统；电压管理；可再生能源发电；双时间尺度；马尔可夫决策过程；深度强化学习

0　引言

随着分布式可再生能源发电在配电系统中渗透率的不断提高，其出力的波动性与不确定性导致系统电压越限频繁[1]。为保证配电系统安全可靠运行，亟需合理的电压管理方法，有效协调具有不同响应特性的调压设备，从而在多时间尺度上挖掘多种设备联合控制电压的潜力，进而抑制可再生能源发电对配电系统电压的影响。

目前，国内外相关学者针对可再生能源高渗透率的配电网电压管理问题进行了广泛研究。其中，基于局部测量信息的下垂控制方法较早应用到分布式电源有功功率和无功功率控制方面，以平抑电压波动[2]，其相应技术规范已写入IEEE1547-2018标准[3]。然而，这种局部调压方法无法协调多种调压设备，较难达到全局电压偏差最小的目标。因此，基于最优潮流（optimal power flow，OPF）模型的电压协调控制得到了广泛研究。其中，基于OPF的电压管理方法可分为2种。1）单时间尺度模型：主要研究短时间尺度（分钟级）分布式电源逆变器等电力电子调压设备的协调控制。文献[4-5]研究了光伏逆变器和静止无功补偿器在分钟级尺度上的协调控制以达到全网节点电压偏差最小的目标。为实现线上实时控制策略，部分研究学者利用梯度映射法[6]、对偶上升法[7]和广义快速对偶上升法[8]等设计了电压反馈控制模型，从而实现基于局部测量信息的全局优化。部分学者进一步研究了该类反馈控制模型的异步迭代机制[9]，以及通信带宽[10]对电压控制的影响。2）双时间尺度模型：为有效协调配电系统中不同时间尺度响应特性的调压设备，提出了双时间尺度的电压模型。文献[11]利用动态规划法与二阶锥规划算法分别求解日前与日内控制策略。文献[12]利用模型预测控制对两阶段电压管理进行建模，并采用分布式算法进行求解。

以上基于优化模型的两类电压管理模型存在以下2个问题。1）潮流方程的凸化问题。在优化问题中一般可通过二阶锥松弛[13-14]、半正定松弛潮流模型[15]、忽略网损的线性化配网潮流模型[16]或电压灵敏度方法[17]在运行点附近的潮流方程进行线性化。以上方法均不可避免带来计算误差。2）随机变量的建模。无论是基于区间数建模进而转化为鲁棒优化问题[18]，还是采取场景法进而转化为随机优化模型[19]或忽略随机变量的方法均不可避免地带来较高的计算复杂度或较大的误差。为解决以上2个问题，本文拟利用深度强化学习方法来求解电压管理的决策问题，通过大量样本的线下计算来达到实时控制的效果。

深度强化学习利用奖励函数对决策行为进行评价，能够根据不同的运行要求和优化目标给出最优动作策略，实现实时决策[20]。已有研究将双深度Q网络（double deep Q network，DDQN）算法[21]、对抗Q算法[22]、rainbow算法[23]分别应用于配电系统的电压管理、需求响应管理和能量调度方面，以解决离散

收稿日期：2021-12-20；修回日期：2022-01-28。
国家自然科学基金资助项目（52107129）。

变量优化问题。文献[24]采用深度确定性策略梯度(deep deterministic policy gradient,DDPG)算法求解包含高维连续优化变量的能量管理问题,回避了对连续变量离散化所导致的误差。文献[25]针对多个光伏逆变器协调控制问题,采用多智能体深度确定性策略梯度(multi-agent DDPG,MADDPG)算法求解,解决了多种调节设备智之间的信息交互问题、有效实现智能体之间协调控制。

在上述背景下,本文针对调压设备不同时间响应特性和动作特性,建立双时间尺度电压管理模型。通过有效考虑两者的时间耦合关系,将其等价转化为马尔可夫决策模型,进而采用深度强化学习算法求解。马尔可夫决策过程通过状态随机转移有效考虑了电压管理模型中的不确定性变量。此外,针对具体的物理模型和控制变量类型,本文分别提出利用DDQN算法求解长时间尺度模型的离散控制变量,利用MADDPG算法求解短时间尺度模型多逆变器的连续控制变量,达到算法和物理模型的有机统一。最后,本文基于IEEE 33节点测试系统建立双时间尺度电压管理模型,通过算例验证本文模型和算法具有更优异的求解精度和鲁棒性。

1 双时间尺度电压管理模型与算法

1.1 双时间尺度电压管理模型

有载调压变压器(on-load tap changer,OLTC)和电容器组(capacitor banks,CB)的动作依靠机械设备,其动作时间尺度为小时级。可再生能源发电设备中的逆变器具有响应速度快、可连续动作等特点。本文设定短时间尺度(分钟级)的控制对象为光伏与风机的逆变器,长时间尺度(小时级)控制对象为OLTC与CB。t时段的建模和求解过程如下所示:

1)长时间尺度的电压管理模型:在优化区间内(本文设置优化区间为1 d,并以小时为时间单位,即时段$T=24$)求解OLTC与CB在t时段的控制策略。优化目标为最小化所有时段每个节点电压与给定值之差的绝对值之和。

2)短时间尺度的电压管理模型:将t时段等分为K个子时段(本文设每时段为1 min,故$K=60$)。基于t时段OLTC与CB的状态求解第τ个子时段内逆变器的控制策略,优化目标为最小化每个节点电压与给定值之差的绝对值之和。

在训练阶段,首先训练短时间尺度神经网络,长时间尺度神经网络的训练需要借助短时间尺度模型来计算其奖励。在应用阶段,长时间尺度模型先做决策,获得OLTC与CB调度策略。在该时段内每一时段,短时间尺度模型根据OLTC与CB状态进行决策,实现电压实时管理。

1.2 双时间尺度电压管理模型算法

目前,深度强化学习算法可分为可以分为2类。一类可以求解离散动作空间问题,如DQN算法、DDQN算法、rainbow算法等;另一类可以解决连续动作空间问题,如DDPG算法、MADDPG算法等。

DDQN算法引入了经验回放机制和2个结构完全相同的估计和目标神经网络,避免了单个神经网络同时进行动作选择与目标Q值计算而造成的过估计问题。因此,本文选择以DDQN算法有效求解OLTC与CB在小时级与离散动作空间下的响应策略以实现长时间尺度电压管理。

与DDPG算法相比,MADDPG算法采用独立采样、统一学习的方法,不仅可降低集中训练的复杂度,还可以很好解决多智能体协作的任务。因此,本文选择MADDPG算法求解光伏与风机智能体逆变器在连续空间上的瞬时协调控制策略,实现短时间尺度的电压管理。

2 基于深度强化学习算法的电压管理模型

配电系统的运行状态可以通过潮流计算获得。通过求解潮流模型可产生大量系统数据以供训练。本文采用节点注入潮流模型,如式(1)—式(3)所示。

$$\begin{cases} P_i = v_i \sum\limits_{ij\in\Omega} v_j \left(Z_{ij}\cos\delta_{ij} + B_{ij}\sin\delta_{ij} \right) \\ Q_i = v_i \sum\limits_{ij\in\Omega} v_j \left(Z_{ij}\cos\delta_{ij} - B_{ij}\sin\delta_{ij} \right) \end{cases} \tag{1}$$

$$P_i = P_{\mathrm{DRES},i} - P_{\mathrm{L},i} \tag{2}$$

$$Q_i = Q_{\mathrm{DRES},i} + Q_{\mathrm{CBs},i} - Q_{\mathrm{L},i} \tag{3}$$

式中:Ω为配网中所有支路的集合;i和j分别为支路ij的输入节点和输出节点;P_{ij}和Q_{ij}分别为支路ij的有功功率和无功功率;P_i和Q_i分别为节点i注入的有功功率和无功功率;Z_{ij}和B_{ij}分别为支路ij的电导和电纳;δ_{ij}为节点i与j之间的相角差;$P_{\mathrm{DRES},i}$为节点i处可再生能源发电机组注入的有功功率;$Q_{\mathrm{DRES},i}$和$Q_{\mathrm{CBs},i}$分别为节点i处可再生能源发电机组和CB注入的无功功率;$P_{\mathrm{L},i}$和$Q_{\mathrm{L},i}$分别为节点i的负荷的有功功率和无功功率;v_i为节点i的电压幅值。

2.1 基于MADDPG算法的短时间尺度电压管理模型

在短时间尺度电压管理模型基于t时段OLTC和CB的状态,求解逆变器在每个时段的最优控制策略。t时段中第τ个子时段的电压管理模型可描述为:

$$\min \sum_{i=1}^{M}|v_{t,i,\tau}(q^g(t,i,\tau))-v_0| \tag{4}$$

s.t. 式(1)—式(3)

$$q^g(t,i,\tau)=Q_{\mathrm{DRES},t,i,\tau} \tag{5}$$

$$|q^g(t,i,\tau)|<\bar{Q}_{\mathrm{DRES},t,i,\tau} \tag{6}$$

$$\underline{v}<v_{t,i,\tau}<\bar{v} \tag{7}$$

$$Q_{\mathrm{DRES},t,i,\tau}=Q_{\mathrm{PV},t,i,\tau}+Q_{\mathrm{WT},t,i,\tau} \tag{8}$$

式中：M为配电系统中的节点数；$v_{t,i,\tau}(q^g(t,i,\tau))$为$t$时段中第$\tau$个子时段节点$i$的电压幅值；$q^g(t,i,\tau)$为$t$时段中第$\tau$个子时段的无功功率；$v_0$为目标电压幅值；$Q_{\mathrm{DRES},t,i,\tau}$和$\bar{Q}_{\mathrm{DRES},t,i,\tau}$为$t$时段中第$\tau$个子时段节点$i$处可再生能源发电装置输出的无功功率及其上限；$Q_{\mathrm{PV},t,i,\tau}$和$Q_{\mathrm{WT},t,i,\tau}$分别为$t$时段中第$\tau$个子时段节点$i$处光伏和风机逆变器所产生的无功功率；$\bar{v}$和$\underline{v}$分别为节点$i$电压幅值的上限和下限。

上述电压管理模型可建模为马尔可夫决策过程，t时段中第τ个子时段的马尔可夫决策过程具体描述如下：

1)短时间尺度状态空间S_{d}：状态空间应包括学习环境中所有智能体的观测状态。每个智能体的状态信息包括当前智能体的出力情况与上一时段调压设备的状态。

2)短时间尺度动作A_{d}：动作空间为学习环境中多个智能体的可以采用的动作。τ时段光伏智能体和风机智能体动作为其各自的无功功率。

3)短时间尺度奖励R_{d}：将优化目标与约束条件映射为奖励函数。以光伏智能体为例，τ时段光伏智能体奖励$R_{\mathrm{PV},t,\tau}$应包括τ时段电压偏差(即优化模型的目标函数)与系统运行约束条件的违规惩罚，可表示为：

$$R_{\mathrm{PV},t,\tau}=\sum_{i=1}^{M}\zeta_0|v_{t,i,\tau}-v_0|+\Gamma_{\mathrm{PV},1,t}+\Gamma_{\mathrm{PV},2,t} \tag{9}$$

$$\Gamma_{\mathrm{PV},1,t}=\begin{cases}0 & |q^g(t,i,\tau)|\leqslant\bar{Q}_{\mathrm{DRES},t,i,\tau}\\ -10\,000 & |q^g(t,i,\tau)|>\bar{Q}_{\mathrm{DRES},t,i,\tau}\end{cases} \tag{10}$$

$$\Gamma_{\mathrm{PV},2,t}=\sum_{i=1}^{M}\zeta_1(|v_{t,i,\tau}-\bar{v}|^{+}+|\underline{v}-v_{t,i,\tau}|^{+}) \tag{11}$$

式中：ζ_0为节点电压偏离基准电压的惩罚系数；$|\cdot|^{+}$为取正函数；ζ_1为节点电压超过给定上限和下限的惩罚系数，且满足$|\zeta_1|\gg|\zeta_0|$；$\Gamma_{\mathrm{PV},1,t}$和$\Gamma_{\mathrm{PV},2,t}$分别为约束式(6)和式(7)的违规惩罚参数。

光伏智能体和风机智能体相互合作以控制全局电压。因而，其两者奖励函数具有相同的形式，根据式(9)—式(11)可得风机智能体的奖励函数。

2.2　基于DDQN算法的长时间尺度电压管理模型

长时间尺度电压管理模型中OLTC与CB的控制策略需考虑优化时段内可再生能源出力的不确定性。因此，假定在t时段内光伏与风机出力功率为随机变量，可将长时间尺度电压管理描述为随机优化模型：

$$\min\ E\Big(\sum_{t=1}^{T}\sum_{i=1}^{M}\sum_{\tau=1}^{K}|v_{t,i,\tau}(q^g(t,i,\tau))-v_0|\Big) \tag{12}$$

s.t. 式(1)—式(3)，式(5)—式(8)

$$Q_{\mathrm{CBs},t,i}=C_{\mathrm{CBs},t,i}q_{\mathrm{CBs},i} \tag{13}$$

$$0\leqslant C_{\mathrm{CBs},t,i}\leqslant\bar{N}_{\mathrm{CBs},i} \tag{14}$$

$$v_{\mathrm{c},t}=h_t v_{\mathrm{p},t} \tag{15}$$

$$h_t=h_0+H_t\Delta h \tag{16}$$

$$-\bar{H}\leqslant H_t\leqslant\bar{H} \tag{17}$$

$$\sum_{t=1}^{T}|H_t-H_{t-1}|\leqslant\overline{A}_{\mathrm{OLTC}} \tag{18}$$

$$\sum_{t=1}^{T}|C_{\mathrm{CBs},t,i}-C_{\mathrm{CBs},t-1,i}|\leqslant\bar{A}_{\mathrm{CBs}} \tag{19}$$

式中：$E(\cdot)$为数学期望函数；$C_{\mathrm{CBs},t,i}$和$\bar{N}_{\mathrm{CBs},i}$分别为t时段内节点i处的CB的投切组数和总组数；$q_{\mathrm{CBs},i}$为节点i处的CB的单位组数的无功功率；$\bar{A}_{\mathrm{CBs}}$为CB优化周期动作次数上限；$v_{\mathrm{c},t}$为t时段内初始时段安装OLTC节点1的电压，一般为固定值1；$v_{\mathrm{p},t}$为t时段平衡节点的电压，用于节点的电压计算；Δh和h_0分别为OLTC的相邻档位调节变化量和初始电压调整率；$\overline{A}_{\mathrm{OLTC}}$为OLTC优化周期动作次数上限；$\bar{H}$为OLTC档位的最大值；$H_t$和$h_t$分别为$t$时段OLTC的档位和电压调整率。

将上述随机优化模型建模为马尔可夫决策过程，如下所示：

1)长时间尺度状态空间S_{c}：在长时间尺度中需要采用期望和标准差来刻画t时段内分布式可再生能源出力情况。因此，t时段的状态包括可再生能源的发电情况、上一时段OLTC与CB的状态、可再生能源的期望值与标准差和调压设备在周期内的动作次数。

2)长时间尺度动作空间A_{c}：长时间尺度的动作空间应包含长时间尺度调压设备OLTC和CB的动作。

3)长时间尺度奖励R_{c}：当t时段不满足一个周期内OLTC与CB的动作次数约束时奖励为一个绝对值很大的负数ϕ；反之，其为t时段中第K个子时段，M个节点的电压偏差之和与K个子时段惩罚之和的总和，表达式为：

$$R_{c,t}=\sum_{\tau=1}^{K}\sum_{i=1}^{M}\zeta_0|v_{t,i,\tau}-v_0|+\sum_{\tau=1}^{K}(\Gamma_{PV,t,\tau}+\Gamma_{WT,t,\tau}) \tag{20}$$

式中：$\Gamma_{PV,t,\tau}$和$\Gamma_{WT,t,\tau}$分别为t时段中第τ个子时段的光伏和风机智能体的约束违规惩罚，其中，$\Gamma_{PV,t,\tau}=\Gamma_{PV,1,t}+\Gamma_{PV,2,t}$。

2.3 电压管理深度网络训练方法

在实际应用中，短时间尺度的逆变器调度策略属于在线优化，可实时读取当前状态，调用MADDPG算法演员网络，获得实时调度策略。长时间尺度（小时级）的调度策略为离线优化，调用DDQN算法目标Q网络，获取长时间尺度调度策略。基于此，长时间尺度基于可再生能源发电出力预测值，获取一天的状态，并调用目标Q网络获取调度策略。双时间尺度电压管理模型中的DDQN算法与MADDPG算法的神经网络更新过程如附录A所示，双时间尺度模型如附录A图A2所示。

3 算例与结果

3.1 参数设置

以IEEE 33节点标准配电系统为基础构造测试系统，对所提方法的可行性与有效性进行验证。系统的基准电压为12.66 kV，在标准系统中分别接入4组光伏和2台风机，其接入位置如图1所示，容量如附录B所示。其中，各节点的负荷数据根据原始IEEE 33节点标准配电系统的负荷数据进行等比例分配。OLTC安装在节点1与节点2之间，共有11个调节位置，每个抽头的调节率为1%。节点33处安装有5组电容器，每组的无功功率为60 kvar。考虑实际运行需要，设置OLTC与CB优化周期内动作次数上限分别为4和5。

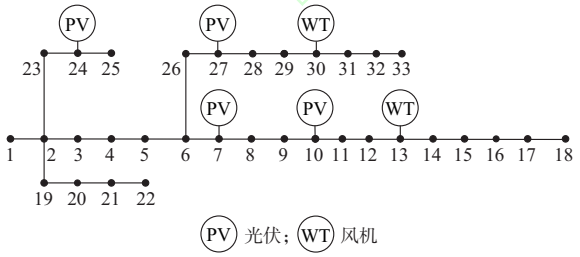


图1 IEEE 33节点系统结构图
Fig. 1 Structure of IEEE 33-bus system

本文在长时间尺度的电压管理模型中考虑了光伏与风机出力的不确定性。假设光伏和风机出力随机变量的分布由预测值和预测误差构成。光伏与风机出力的期望为其预测值。假设光伏出力预测的方差为其期望的3%，风机出力预测的方差为其期望的5%。随机抽样生成1 020组场景，其中，1 000组场景用于后文随机优化模型的策略求解，20组对所得策略进行验证。此外，非线性潮流方程通过调用Pypower包求解。

3.2 训练过程

所提模型的求解算法主要包括DDQN算法与MADDPG算法2个部分。

MADDPG算法中设置光伏和风机逆变器两类智能体。每个智能体的演员网络为策略选择网络，当该网络得出的策略奖励值趋于稳定时，代表该智能体深度网络收敛。训练过程中，设定演员网络的学习率为0.001，评论家网络的学习率为0.002。衰减因子γ为0.9；小批量采样规模为100，样本池内样本总量为10 000，观测步数1 800。电压违规惩罚因子ζ_1与ζ_0分别为−1 000和−2。设定动作探索噪声随着训练步数逐渐减小，初始噪声服从$N(0,1)$的正态分布。当训练步数大于观测步数时，噪声服从$N(0,0.995^z)$的正态分布，其中，z为开始训练的步数。训练过程如图2所示。

本文以DDQN算法求解长时间尺度OLTC与CB的控制策略。在DDQN神经网络训练过程中，本文设定经验回放机制的样本存储量为10 000，每次小批量采样规模为100；初始探索率为0.1，最终探索率为0.001，探索步数为5 000；学习率取0.001；每训练10次更新一次目标Q网络参数，惩罚因子ϕ为−100 000。训练过程如图2(c)所示。

由图2(a)和(b)可知，光伏智能体与风机智能体的演员网络分别在训练4 000步和3 500步时达到了收敛状态。随着训练步数的增加，2类智能体演员网络的奖励分别收敛于−0.632与−0.562。由图2(c)可得，在12 000步时DDQN算法神经网络达到收敛状态。在与环境的不断交互学习中，奖励值最终收敛于−1.212。最优策略的奖励收敛于−1.212而非奖励最大值为−1.205，其差值可看作可再生电源出力不确定性所带来的风险成本。

3.3 算例结果与分析

为进一步研究所提出的模型与方法的有效性，本部分设置了3种模式进行对比分析，具体如下：

模式0：不进行任何优化管理。

模式1：基于随机优化算法求解双时间尺度配电系统电压管理策略，模型见文献[9]。

模式2：本文所提算法，基于DDQN算法求解长时间尺度电压管理模型，基于MADDPG算法求解短时间尺度的电压管理模型。

采用3.1节随机生成的20个场景对上述3种模

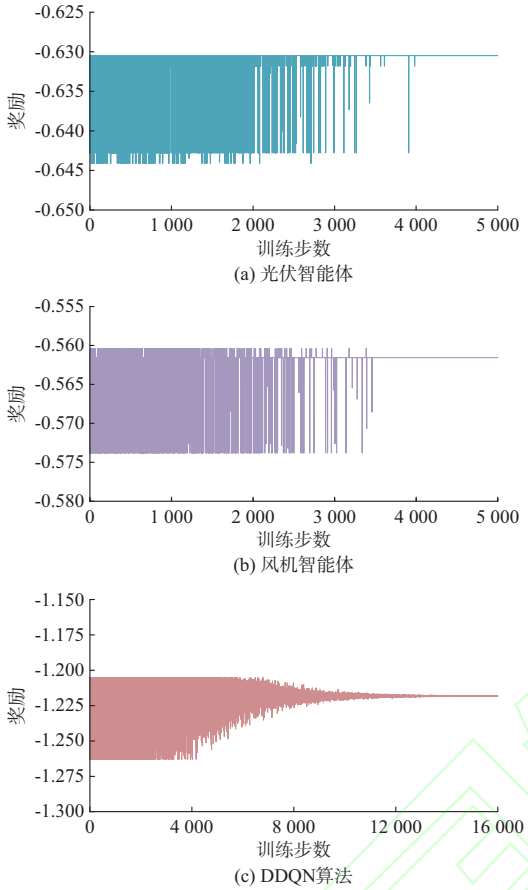


图2　深度强化学习算法的神经网络训练过程
Fig. 2　Neural network training process of deep reinforcement learning algorithm

式所得策略进行验证。3种模式的测试结果如表1所示。违规时段数表示优化周期内24个时段中存在电压违规现象的时段个数；平均电压偏差为33个节点24个时段电压偏差的平均值；电压最大爬坡和最小爬坡分别为各节点24个时段最大电压变化和最小电压变化。

表1　优化结果对比
Table 1　Comparison of optimization results

模式	违规时段数/个	平均电压偏差/p.u.	最大爬坡/p.u.	最小爬坡/p.u.	计算时间/s
模式0	4	0.031 8	0.060	0.000 6	
模式1	0	0.018 7	0.044	0.000 6	34.25
模式2	0	0.013 8	0.062	0.000 2	0

从表1可知，模式1和模式2所得策略均不存在电压违规现象。与模式1相比，模式2优化后的电压偏差更小，仅为0.013 8，并可通过训练好的神经网络实时决策得到。

为更加直观地对比模式1与模式2的优化结果，以模式0中存在电压越限的07:00和20:00时的数据为例，展示3种模式所得各节点电压，如图3(a)和(b)所示。由图3可知，模式1与模式2都能有效控制配电系统电压运行在安全范围内。与模式1相比，模式2的优化结果更加靠近基准电压值且电压幅值波动更加平缓，因而本文所提算法在电压偏差控制上表现更好。

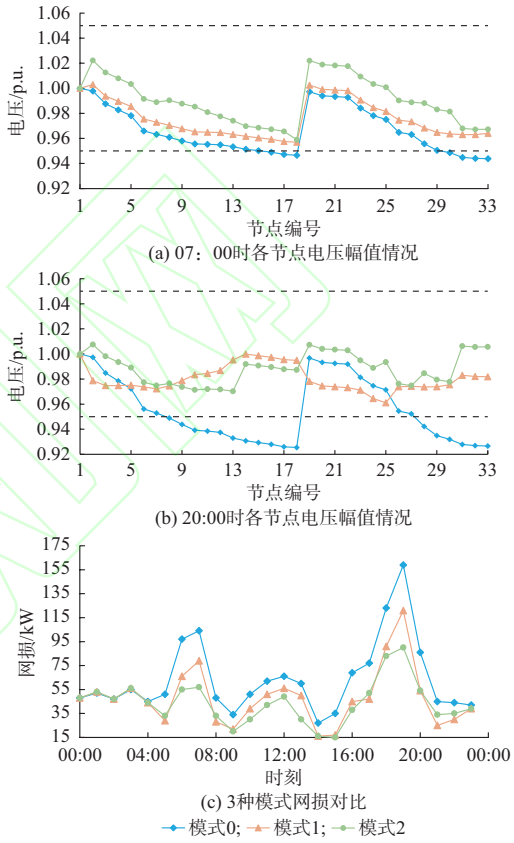


图3　3种模式电压优化结果对比
Fig. 3　Comparison of voltage optimization results in three modes

为进一步分析所提方法的优越性，拟对比各种运行策略下的系统网损，对比结果如图3(c)所示。由图3(c)可知，本文所提方法在3种模式中系统的网损最低。与模式1相比，模式2的平均网损降低了8.67%，这是因为模式2下节点电压靠近基准电压，且其幅值变化平缓，而网损与相邻节点之间的电压差成正相关，一定程度上降低了系统网损。由此可知，本文所提策略在保证配电系统安全运行的基础上，可间接有效提高配电系统运行的经济性。

3.4　算法鲁棒性分析

本部分设置极端场景以验证上述模式1和2所得策略的鲁棒性。假设在t时段中，每个时段可再生能源的出力满足的条件为：

$$\begin{cases} P_{\mathrm{DRES},t,\tau} \leqslant \tilde{P}_{\mathrm{DRES},t} - 2\sigma_{\mathrm{DRES},t} & \tau\text{为奇数} \\ P_{\mathrm{DRES},t,\tau} \geqslant \tilde{P}_{\mathrm{DRES},t} + 2\sigma_{\mathrm{DRES},t} & \tau\text{为偶数} \end{cases} \tag{21}$$

式中：$\tilde{P}_{\mathrm{DRES},t}$和$\sigma_{\mathrm{DRES},t}$为可再生能源$t$时段出力的期望值与标准差；$P_{\mathrm{DRES},t,\tau}$为$t$时段中第$\tau$个子时段分布式可再生能源的有功功率。

在此极端场景下，验证2种模式下电压偏差情况。通过仿真模拟，模式1的平均电压偏差为0.021 9，模式2下平均电压偏差为0.014 2。因此，模式2在面对不确定性环境具有更好的鲁棒性。

相对于DDPG算法，MADDPG算法可以有效解决多个智能体的交互问题，并具有鲁棒性。为进一步分析所提MADDPG算法的鲁棒性，本文设置模式3：

模式3：基于DDQN算法求解长时间尺度配电系统电压管理策略，并基于DDPG算法求解短时间尺度的配电系统电压管理策略。

DDPG算法神经网络训练过程中，设定演员网络的学习率，评论家网络的学习率、衰减因子γ、目标演员网络与目标评论家网络的更新因子ε与MADDPG算法相同；小批量采样的规模为80，样本池内样本总量为10 000，观测步数1 500，动作探索噪声随着训练步数逐渐减小，初始噪声服从$N(0,2)$的正态分布。

为验证MADDPG算法的鲁棒性，设置4种光伏故障情况来测试模式2和模式3所得策略的性能。假设光伏智能体链路故障时无法提供有功和无功功率支撑。故障情况具体如下：

故障0：无故障发生。

故障1：节点24的光伏机组故障。

故障2：节点24、节点10的光伏机组故障。

故障3：节点24、节点1、节点7的光伏机组故障。

针对上述4种光伏机组故障比较DDPG算法与MADDPG算法的性能，结果如表2所示。

表2　4种故障下的优化结果对比
Table 2　Comparison of optimization results in four types of faults

故障	违规时段数/个			平均电压偏差/p.u.		
	模式0	模式2	模式3	模式0	模式2	模式3
故障0	4	0	0	0.031 81	0.013 80	0.013 22
故障1	4	0	0	0.032 07	0.013 61	0.013 25
故障2	4	0	0	0.032 11	0.013 83	0.015 85
故障3	5	0	1	0.032 19	0.013 89	0.030 68

由表2可知，在故障0与故障1的情况下，模式2与模式3的所得电压偏差较小。在故障2与故障3的情况下，由于光伏智能体信息数据丢失严重，与模式2相比，模式3不能很好地处理，出现了电压违规现象，且平均电压偏差较大。由此可知，MADDPG算法在面对光伏机组故障时具有更好的鲁棒性。

4　结语

本文针对配网内调压设备不同时间响应特性，建立双时间尺度的电压管理模型。在有效考虑可再生能源出力不确定性的基础上，本文将电压管理模型等价转化为马尔可夫决策模型。针对多种电压控制对象的动作与响应等技术特点，分别利用深度强化学习DDQN算法和MADDPG算法分别求解长时间尺度模型的离散控制变量和短时间尺度模型的连续控制变量，实现了配电系统电压管理的实时决策。与传统随机优化算法相比，本文所提方法可得到更优异的电压管理效果，同时在面对随机环境时具有更好的鲁棒性。

在后续研究中，拟采用随机过程对电压管理模型的随机变量进行建模，通过考虑时序上的关联性来降低本文策略的保守性。

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] 卓振宇，张宁，谢小荣，等. 高比例可再生能源电力系统关键技术及发展挑战[J]. 电力系统自动化，2021，45(9)：171-191.
ZHUO Zhenyu, ZHANG Ning, XIE Xiaorong, et al. Key technologies and developing challenges of power system with high proportion of renewable energy[J]. Automation of Electric Power Systems, 2021, 45(9): 171-191.

[2] 杨美辉，周念成，王强钢，等. 基于分布式协同的双极直流微电网不平衡电压控制策略[J]. 电工技术学报，2021，36(3)：634-645.
YANG Meihui, ZHOU Niancheng, WANG Qianggang, et al. Unbalanced voltage control strategy of bipolar DC microgrid based on distributed cooperation [J]. Transactions of China Electrotechnical Society, 2021, 36(3): 634-645.

[3] Interconnection and interoperability of distributed energy resources with associated electric power systems interfaces: IEEE Standard 1547—2018 [S/OL]. [2021-12-20]. https://ieeexplore.ieee.org/document/8332112.

[4] FENG C S, LI Z Y, MOHAMMAD S, et al. Decentralized short-term voltage control in active power distribution systems [J]. IEEE Transactions on Smart Grid, 2018, 9(5): 4566-4576.

[5] 张宏，董海鹰，陈钊，等. 基于模型预测控制的光热-光伏系统多

时间尺度无功优化控制策略研究[J].电力系统保护与控制，2020,48(9):135-142.
ZHANG Hong, DONG Haiying, CHEN Zhao, et al. Multi-time scale reactive power optimal control strategy of a CSP-PV system based on model predictive control [J]. Power System Protection and Control, 2020, 48(9): 135-142.

[6] ZHU H, LIU H J. Fast local voltage control under limited reactive power: optimality and stability analysis [J]. IEEE Transactions on Power Systems, 2016, 31(5): 3794-3803.

[7] LI J Y, XU Z, ZHAO J, et al. Distributed online voltage control in active distribution networks considering PV curtailment [J]. IEEE Trans Industrial Informatics, 2019, 15(10): 5519-5530.

[8] GUO Y F, GAO H L, WANG Z Y. Distributed online voltage control for wind farms using generalized fast dual ascent [J]. IEEE Transactions on Power Systems, 2020, 35(6): 4505-4517.

[9] SINDRI M, QU G N, LI N. Distributed optimal voltage control with asynchronous and delayed communication [J]. IEEE Transactions on Smart Grid, 2020, 11(4): 3469-3482.

[10] DING T, YANG Q R, YANG Y H, et al. A data-driven stochastic reactive power optimization considering uncertainties in active distribution networks and decomposition method [J]. IEEE Transactions on Smart Grid, 2018, 9(5): 4994-5004.

[11] 褚国伟，张友旺，葛乐，等.自储能柔性互联配电网多时间尺度电压优化[J].电力系统自动化，2021,45(9):71-79.
CHU Guowei, ZHANG Youwang, GE Le, et al. Multi-time-scale voltage optimization of flexible interconnected distribution network with self-energy storage [J]. Automation of Electric Power Systems, 2021, 45(9): 71-79.

[12] GUO Y F, WU Q W, GAO H L, et al. Double-time-scale coordinated voltage control in active distribution networks based on MPC [J]. IEEE Transactions on Sustainable Energy, 2020, 11(1): 294-303.

[13] 姜涛，张东辉，李雪，等.含分布式光伏的主动配电网电压分布式优化控制[J].电力自动化设备，2021,41(9):102-109.
JIANG Tao, ZHANG Donghui, LI Xue, et al. Distributed optimal control of voltage in active distribution network with distributed photovoltaic [J]. Electric Power Automation Equipment, 2021, 41(9): 102-109.

[14] 陈瑞捷，鲁宗相，乔颖.基于多场景模糊集和改进二阶锥方法的配电网优化调度[J].电网技术，2021,45(12):4621-4629.
CHEN Ruijie, LU Zongxiang, QIAO Ying. Optimal dispatch based on multi-scene ambiguity set and modified second-order cone algorithm for distribution network [J]. Power System Technology, 2021, 45(12): 4621-4629.

[15] 张亚超，黄张浩，郑峰，等.基于风电出力模糊集的电-气耦合系统分布鲁棒优化调度[J].电力系统自动化，2020,44(4):44-53.
ZHANG Yachao, HUANG Zhanghao, ZHENG Feng, et al. Distributionally robust optimal dispatch for power-gas coupled system based on fuzzy set of wind power output[J]. Automation of Electric Power Systems, 2020, 44(4): 44-53.

[16] YANG Q L, WANG G, ALIREZA S, et al. Two-timescale voltage control in distribution grids using deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2019, 11(3): 2313-2323.

[17] 彭琰，余一平，鞠平，等.计及不确定性的电力系统电压波动分析方法[J].电力自动化设备，2017,37(8):137-142.
PENG Yan, YU Yiping, JU Ping, et al. Voltage fluctuation analysis method considering uncertainties of power system [J]. Electric Power Automation Equipment, 2017, 37(8): 137-142.

[18] 张亚超，易杨，胡志鹏，等.基于分布鲁棒优化的电-气综合能源系统弹性提升策略[J].电力系统自动化，2021,45(13):76-84.
ZHANG Yachao, YI Yang, HU Zhipeng, et al. Resilience enhancement strategy of electricity-gas integrated energy system based on distributionally robust optimization [J]. Automation of Electric Power Systems, 2021, 45(13): 76-84.

[19] 刘玉奇，臧传治，王悦，等.基于随机经济模型预测控制的电热综合能源系统运行优化[J].电力自动化设备，2021,41(7):14-21.
LIU Yuqi, ZANG Chuanzhi, WANG Yue, et al. Optimal operation of electricity-heating integrated energy system based on stochastic economic model predictive control [J]. Electric Power Automation Equipment, 2021, 41(7): 14-21.

[20] 杨挺，赵黎媛，刘亚闯，等.基于深度强化学习的综合能源系统动态经济调度[J].电力系统自动化，2021,45(5):39-47.
YANG Ting, ZHAO Liyuan, LIU Yachuang, et al. Dynamic economic dispatch for integrated energy system based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2021, 45(5): 39-47.

[21] 王甜婧，汤涌，郭强，等.基于知识经验和深度强化学习的大电网潮流计算收敛自动调整方法[J].中国电机工程学报，2020,40(8):2396-2406.
WANG Tianjing, TANG Yong, GUO Qiang, et al. Automatic adjustment method of power flow calculation convergence for large-scale power grid based on knowledge experience and deep reinforcement learning [J]. Proceedings of the CSEE, 2020, 40(8): 2396-2406.

[22] 于一潇，杨佳峻，杨明，等.基于深度强化学习的风电场储能系统预测决策一体化调度[J].电力系统自动化，2021,45(1):132-140.
YU Yixiao, YANG Jiajun, YANG Ming, et al. Prediction and decision integrated scheduling of energy storage system in wind farm based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2021, 45(1): 132-140.

[23] WANG B, LI Y, MING W Y, et al. Deep reinforcement learning method for demand response management of interruptible load [J]. IEEE Transactions on Smart Grid, 2020, 11(4): 3146-3155.

[24] 彭刘阳，孙元章，徐箭，等.基于深度强化学习的自适应不确定性经济调度[J].电力系统自动化，2020,44(9):33-42.
PENG Liuyang, SUN Yuanzhang, XU Jian, et al. Self-adaptive uncertainty economic dispatch based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2020, 44(9): 33-42.

[25] CAO D, HU W H, ZHAO J B. A multi-agent deep reinforcement learning based voltage regulation using coordinated PV inverters [J]. IEEE Transactions on Power

Systems, 2020, 35(5): 4120-4123.

冯昌森(1990—),男,博士,讲师,硕士生导师,主要研究方向:高电力系统优化控制、电力市场、机器学习。E-mail:fcs@zjut.edu.cn

张　瑜(1996—),女,硕士研究生,主要研究方向:电力系统优化控制、机器学习。E-mail:1304150284@qq.com

张有兵(1971—),男,通信作者,教授,博士生导师,主要研究方向:智能电网、分布式发电及新能源优化控制、电动汽车入网、电力系统通信、电能质量监控。E-mail:youbingzhang@zjut.edu.cn

(编辑　杨松迎)

Deep Reinforcement Learning Approach for Dual-Timescale Voltage Management in Distribution System

FENG Changsen[1], *ZHANG Yu*[1], *XIE Luyao*[1], *WEN Fushuan*[2], *ZHANG Kaiyi*[1], *ZHANG Youbing*[1]

(1. College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023 , China;
2. School of Electrical Engineering, Zhejiang University, Hangzhou 310027, China)

Abstract: With the increasing penetration of renewable energy generation, the problem of voltage violation in the distribution system becomes more frequent, and efficient voltage management strategies are urgently needed to ensure the secure and economic operation of the distribution system. First, this paper establishes a dual-timescale voltage management model for the distribution system to realize coordinated control of voltage regulators with different time response characteristics. Then, the voltage management models of the two time scales are modeled as Markov decision processes (MDP). Based on effectively considering the temporal coupling relationship between the two and the physical characteristics of controllable devices, the dual-timescale real-time voltage management is realized by using the multi-agent depth deterministic policy gradient algorithm and the double deep Q network algorithm to solve the model, respectively. Finally, the effectiveness of the proposed method is demonstrated by case studies on the IEEE 33-bus standard distribution system.

This work is supported by National Natural Science Foundation of China (No. 52107129).

Key words: distribution system; voltage management; renewable energy generation; two-timescale; Markov decision process; deep reinforcement learning

附录A

DDQN算法在估计Q网络获得OLTC与CB的控制策略为：

$$a=\arg\max_{a\in A} q(s,a;\theta_e) \tag{A1}$$

式中：θ_e为估计Q网络的参数；$q(s,a;\theta_e)$为状态s和下选择动作a时估计Q网络的输出值。

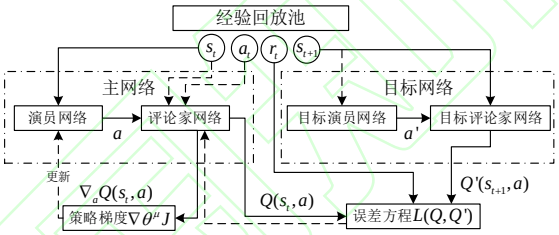
DDQN算法为获得最优控制策略，需在经验回放池中小批次抽样和学习，即通过最小化估计Q网络输出的q值y_g与目标q值$q(s,a;\theta_t)$的误差L，进行训练并更新估计Q神经网络参数为：

$$L=\frac{1}{G}\sum_{g=1}^{G}(y_g-q_g(s,a;\theta_t))^2 \tag{A2}$$

$$q(s,a;\theta_t)=r(s,a)+\gamma q(s',a;\theta_t) \tag{A3}$$

式中：G为训练神经网络时的抽样样本数量；θ_t为目标Q网络的参数；$r(s,a)$为状态s下选择动作a的奖励；γ为折扣因子。

MADDPG算法与DDPG算法的神经网络训练过程相同。每个智能体包括4个神经网络，分别为演员网络、评论家网络、目标演员网络和目标评论家网络，分别用θ^{μ}、θ^{Q}、$\theta^{\mu'}$和$\theta^{Q'}$表示。每个智能体采用DDPG算法进行学习，其原理如图A1所示。



图A1　DDPG算法原理图

Fig. A1　Schematic diagram of DDPG algorithm

以光伏智能体神经网络的学习为例进行说明。光伏智能体评论家网络参数的更新方式与DDQN算法估计Q网络参数更新方式相同，即如式(A2)进行，这里不再赘述。

输入当前状态$s_{\mathrm{PV},t,\tau}$，光伏智能体根据演员网络和探索噪声选择得出动作，如式(A4)和式(A5)所示。

$$a_{\mathrm{PV},t,\tau}=\mu(s_{\mathrm{PV},t,\tau}|\theta^{\mu})+\vartheta_t \tag{A4}$$

$$\mu(s_{\mathrm{PV},t,\tau}|\theta^{\mu})=\arg\max_{\mu}J_{\pi}(\mu|s_{\mathrm{PV},t,\tau},\theta^{\mu}) \tag{A5}$$

式中：ϑ_t为探索噪声。

光伏智能体神经网络在训练过程中需借助风机智能体的动作信息来辅助训练，进而实现多智能体的交互协作。因而光伏智能体评论家网络的动作输入为光伏智能体动作与风机智能体动作的集合。

光伏智能体根据训练得到的控制策略梯度$\nabla_{\theta^{\mu}}J$来更新演员网络参数θ^{μ}，公式如下：

$$\nabla_{\theta^{\mu}}J\approx\frac{1}{W}\sum_{w=1}^{W}\nabla_a q(s,a|\theta^{Q})|_{s=s_w,a=\mu(s_w)}\nabla_{\theta^{\mu}}\mu(\theta^{\mu})|_{s=s_w} \tag{A6}$$

式中：W为样本数量。

DDQN算法中每训练一定步数，将估计Q网络的参数直接复制给目标Q网络。MADDPG算法目标演员网络和目标评论家网络的参数$\theta^{\mu'}$和$\theta^{Q'}$更新方法与DDQN算法不同。MADDPG算法采用软更新的方式，即分别赋予目标网络和原始网络一定的权重，且每一步都会对目标演员网络与目标评论家网络参数进行更新。更新方式如下：

$$\begin{cases}\theta^{Q'}\leftarrow u\theta^{Q}+(1-u)\theta^{Q'}\\ \theta^{\mu'}\leftarrow u\theta^{\mu}+(1-u)\theta^{\mu'}\end{cases} \tag{A7}$$

式中：u为更新因子。

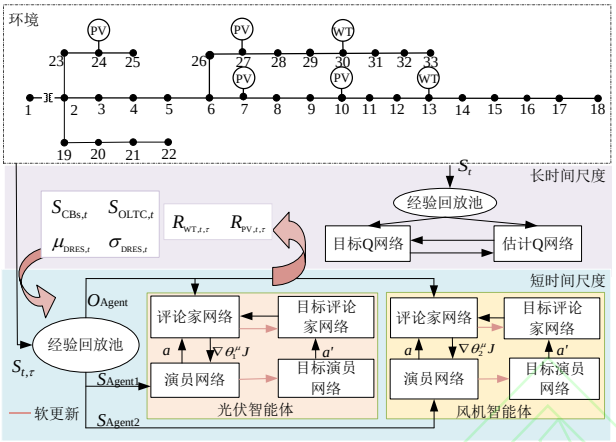


图 A2　双时间尺度电压调节模型原理

Fig. A2　Principle of duel-timescale voltage regulation model

附录 B

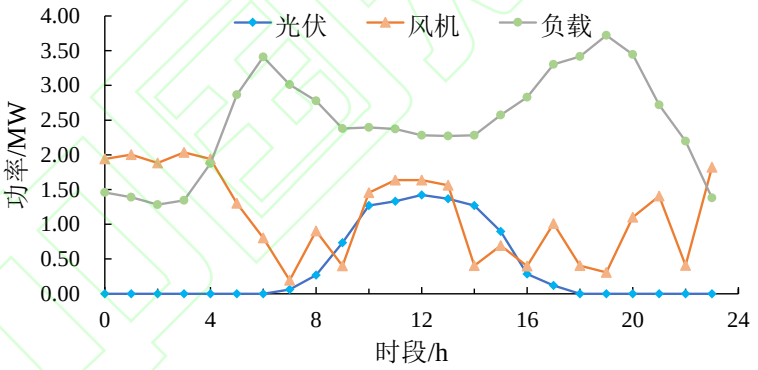


图 B1　可再生能源和负荷的日运行曲线

Fig. B1　Daily operation curves of renewable energy generation and load demand

表 B1　可再生能源发电配置参数

Table B1　Configuration parameters of renewable energy generation

DERS类型	接入节点(接入容量/(kV·A))			
光伏	7 (500)	10 (400)	24 (300)	27 (500)
风机	13 (1 000)		30 (1000)	