第46卷 第8期 2022年4月25日
DOI：10.7500/AEPS20210607005
电力系统自动化
Automation of Electric Power Systems
Vol. 46 No. 8 Apr. 25，2022

弹性配电系统动态负荷恢复的深度强化学习方法

黄玉雄，李更丰，张理寅，别朝红
（西安交通大学电气工程学院，陕西省西安市 710049）

摘要： 极端事件下，合理利用微网盈余电力供应配电系统中的关键负荷，可有效提升电网弹性。基于深度强化学习(DRL)技术，提出一种考虑微网参与的配电系统动态关键负荷恢复(DCLR)方法，支持以无模型的方式求解复杂问题，以提升在线计算效率。首先，分析含微网的配电系统DCLR问题，并在此基础上构建其马尔可夫决策过程，其中考虑了配电系统运行、微网运行和用户满意度等约束条件。其次，基于OpenDSS构建DCLR模拟环境，形成DRL应用所需的智能体-环境交互接口，进一步采用深度Q网络算法搜寻关键负荷恢复的最优控制策略，并定义收敛性、决策能力指标分别用于评价智能体的训练和应用表现。最后，基于改进的IEEE测试系统验证了所提方法的有效性。

关键词： 负荷恢复；深度学习；强化学习；配电系统；微网(微电网)；弹性

0 引言

自然灾害、人为攻击和严重技术故障等极端事件严重威胁着电力系统的安全可靠运行[1]。面向极端事件的电力系统弹性研究已越来越受到重视。电力系统弹性是指电力系统承受和抵御极端事件的能力，包括事前预防、事中适应和事后恢复3个主要阶段[2]。随着碳中和的逐步推进，未来配电系统将接入大量分布式电源(distributed generator，DG)，而微网作为配电系统承载DG的重要形式，其相应技术将得到大力发展。当极端事件导致主网故障而出现供电能力不足时，配电系统可以通过操作分段开关实现分区，各分区内基于微网盈余电力可为配电系统用户提供电力服务。

近年来，配电系统的弹性评估及提升已成为研究热点。文献[3]分析了能源转型下的弹性电力系统的内涵和前景。文献[4]提出了配电系统弹性的核心特征与关键技术。文献[5-6]提出了电力系统弹性评估与提升的通用研究框架。文献[7-9]研究了极端天气下电网的脆弱性建模方法与恢复措施。文献[10]对主动配电系统的弹性研究现状进行了系统性的综述。对于极端事件，相比于事前的设备强化投资，事后或事中的高效负荷恢复被认为是更加经济可行的电网弹性提升方案。文献[11]较早提出了考虑微网参与的配电系统事后运行模式，将弹性配电系统的关键负荷恢复问题建模为混合整数线性规划问题。文献[12]提出了多源协同的配电系统多时段负荷恢复优化决策方法。此外，现有研究还考虑了微网频率稳定性[13]、配电网络拓扑重构[14-15]、配电系统线性拓扑约束[16]、储能设备与需求侧管理[17-18]、分层故障管理[19]、可再生DG出力不确定性[20]等因素对含微网的配电系统负荷恢复的影响。

目前，大多数的关键负荷恢复研究是将恢复问题建模为数学规划模型，进而基于最优化理论求解。这类基于模型的方法存下以下问题：1)需要建立研究对象的显式数学模型，这对于复杂系统而言存在很大难度；2)最优化理论中对于可解性的要求使得模型中存在不同程度的简化；3)当在线应用时，优化问题的高效求解仍然面临着挑战。针对上述问题，无模型的机器学习方法被认为是一种可能的解决途径[21]。

本文基于无模型的深度强化学习(deep reinforcement learning，DRL)技术，研究含微网的弹性配电系统在失去主网供电能力情况下的事后动态关键负荷恢复(dynamic critical load restoration，DCLR)问题。这里的“动态”取自动态规划[22]的概念，有别于电力系统动态特性。相较于传统的静态关键负荷恢复模型，本文将恢复问题建模为马尔可夫决策过程(Markov decision process，MDP)，恢复过程被拆分为一系列单阶段问题逐次求解，各阶段

收稿日期：2021-06-07；修回日期：2021-11-08。
上网日期：2022-02-24。
国家电网有限公司总部科技项目(SGJX0000KXJS1900322)。

决策考虑了未来一定步数决策的影响，各阶段的解组成全过程最优的决策序列。具体而言，首先分析考虑微网的DCLR问题，并在此基础上建立相应的MDP模型，同时构建基于OpenDSS[23]的DCLR模拟环境，形成“智能体-环境”交互接口（agent-environment interface，AEI）。然后，采用深度Q网络（deep Q-network，DQN）[24-25]算法寻求最优控制策略，并定义评价指标用于定量评估智能体在训练和应用环节的表现。最后，基于改进的IEEE测试系统进行算例分析，以验证所提方法的有效性。

1　问题建模

首先，介绍MDP的基本概念与DCLR的数学模型，在此基础上，通过定义状态空间、动作空间等MDP要素，建立DCLR问题的MDP模型。最后，基于OpenDSS建立DCLR模拟环境，形成DRL应用所需的智能体-环境交互接口。

1.1　MDP基本概念

MDP从智能体和环境的互动中学习控制策略，以实现控制目标，是序贯决策的经典表现形式，其由五元组$\{S, A, P, r, \gamma\}$表示，其中S为状态空间；A为动作空间；$P(s'|s, a)$为在采取动作a后由状态s到达状态s'的概率；$r(s, a, s')$为在采取动作a后由状态s到达状态s'的立即奖励；γ为折扣因子，$0 \leqslant \gamma \leqslant 1$。

在智能体与环境的交互中，首先，在时刻t，智能体基于环境的状态$s_t(\in S)$选择动作$a_t(\in A)$；然后，环境根据概率$P(s_{t+1}|s_t, a_t)$转移到状态s_{t+1}，智能体获得立即奖励$r_t = r(s_t, a_t, s_{t+1})$。智能体在交互过程中通过学习获得决策能力，决策目标是最大化其处于给定状态s或状态-动作对(s, t)时的价值，这一价值可由累计奖励的期望进行估计。动作价值函数$Q_\pi(s, a)$的数学表达式为：

$$Q_\pi(s, a) = E_{\tau\sim\pi}(R_{T\tau}|s_\tau = s, a_\tau = a) \quad \forall s \in S, a \in A \tag{1}$$

$$R_{T\tau} = \sum_{k=t}^{T-1} \gamma^{k-t} r_k \tag{2}$$

式中：π为控制策略，表示状态与选择每个可能动作概率的映射；$E_{\tau\sim\pi}(\cdot)$为状态s从采取动作a开始遵循策略π所获得的期望折扣回报函数；τ为MDP轨迹；$R_{T\tau}$为从时刻t开始到时刻T结束的MDP片段（episode）的折扣回报。

γ用于描述预期的未来奖励对当前决策的影响，γ越接近0，表明智能体的目标越“近视”（即向前考虑的步数越少），智能体的训练难度越小，反之，若γ越接近1，则智能体向前考虑的步数越多，智能体的训练难度越大。

1.2　动态关键负荷恢复

极端事件下，配电系统可能在一段时期内无法通过变电站或馈线获取主网电力，在此期间，配电系统运营商（distribution system operator，DSO）可以基于分段开关和DG将配电系统分解成多个孤岛微网，从而为关键负荷继续供电，并提升电网弹性[26]。

DCLR的动态决策过程如图1所示。其中，恢复时间被离散化，记为集合Γ，$|\Gamma| = T$。DSO在各时间点$t \in \Gamma$均会进行负荷恢复决策，且DSO的每次决策受到当前和未来奖励的影响。

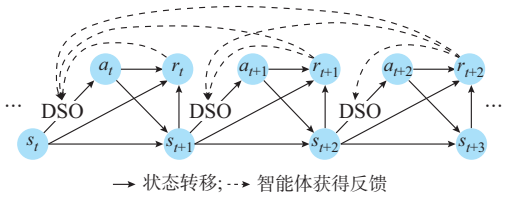

图1　DCLR问题的动态决策示意图

Fig. 1　Schematic diagram of dynamic decision-making for DCLR problem

将配电系统记为图$G = (N, L)$，其中N和L分别表示节点集合和线路集合。微网所在节点的集合记为$M \subset N$。分段开关以集合W表示，每个节点$i(\in N)$的有功功率和无功功率分别记为$p_{t,i,\phi}$和$q_{t,i,\phi}$，其中$\phi \in \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$表示相别；$\omega_i$表示节点$i$负荷的权重系数。以下将从恢复目标和约束条件2个方面进一步分析DCLR问题的数学模型，以此作为后续建立DCLR的MDP以及AEI的基础。

1.2.1　恢复目标

DCLR的目标为在恢复期间的加权负荷恢复总量的数学期望最大，可表示为：

$$\max E_s\left(\sum_{t\in\Gamma, i\in N, \phi\in\{\mathrm{a},\mathrm{b},\mathrm{c}\}} \omega_i \tilde{p}_{t,i,\phi}\right) \tag{3}$$

式中：$E_s(\cdot)$为考虑了系统状态s中随机扰动后的数学期望函数；$\tilde{p}_{t,i,\phi}$为实际恢复负荷，$\tilde{p}_{t,i,\phi} = v_{t,i,\phi} p_{t,i,\phi}$，其中，$v_{t,i,\phi}$为二值状态变量，用于表示负荷$p_{t,i,\phi}$的恢复状态，$v_{t,i,\phi} = 1$表示已被恢复，$v_{t,i,\phi} = 0$表示未被恢复，当微网供电范围确定时，$v_{t,i,\phi}$也随之确定。

1.2.2　约束条件

本文在DCLR问题中考虑3类约束，分别为配电系统运行约束、微网运行约束以及用户满意度约束。

1）配电系统运行约束

配电系统负荷恢复应满足三相不平衡潮流方程：

$$p_{t,i,\phi}-\mathrm{j}q_{t,i,\phi}=(\bar{V}_{t,i,\phi})^{*}\sum_{j\in\Omega_i}\sum_{\phi'}Y_{ij,\phi\phi'}\bar{V}_{t,j,\phi'}$$
$$t\in\Gamma,i\in N,\phi\in\{\mathrm{a},\mathrm{b},\mathrm{c}\},\phi'\in\{\mathrm{a},\mathrm{b},\mathrm{c}\} \quad (4)$$

式中：$\bar{V}_{t,i,\phi}$为时刻t节点i相别ϕ的电压；Ω_i为与节点i直接相连的节点集合；$Y_{ij,\phi\phi'}$为节点导纳矩阵中节点i(相别ϕ)和节点j(相别ϕ')所对应的元素。

配电系统各节点电压应该维持在可行的范围内：

$$V_{\min}\leqslant V_{t,i,\phi}\leqslant V_{\max}\quad t\in\Gamma,i\in N,\phi\in\{\mathrm{a},\mathrm{b},\mathrm{c}\} \quad (5)$$

式中：$V_{t,i,\phi}$为时刻t节点i相别ϕ的电压幅值；$V_{\min}$和$V_{\max}$分别为节点电压幅值的下限和上限，本文取$V_{\min}=0.95$ p.u.，$V_{\max}=1.05$ p.u.。

配电系统各线路电流不应超过其限值：

$$I_{l,\min}\leqslant I_{t,l,\phi}\leqslant I_{l,\max}\quad t\in\Gamma,l\in L,\phi\in\{\mathrm{a},\mathrm{b},\mathrm{c}\} \quad (6)$$

式中：$I_{t,l,\phi}$为时刻t线路l相别ϕ的电流幅值；$I_{l,\min}$和$I_{l,\max}$分别为线路l电流幅值的下限和上限。

2)微网运行约束

在配电系统失去主网供电后，微网处于孤岛运行模式，微网中的DGs不仅可以作为备用机组向本地负荷供电，而且可以通过操作分段开关或微网开关为微网之外的配电系统负荷提供电力服务。假设微网中的DGs采用主从控制策略[27]，其中系统的电压和频率仅由一个DG(即主机)控制，其余DGs(作为从机)工作在电流控制模式下。因此，当配电系统被分解成多个孤立的微网时，各配电系统分区应满足辐射状网络拓扑约束，即每个关键负荷仅由1个微网通过1条路径供电，任意2个微网之间不存在路径(本文不考虑电网簇[28])。实际上，保持辐射状的网络结构有助于简化许多运行问题，例如微网之间的同步和负荷分配问题。此外，辐射状网络使得继电保护装置更加容易整定，系统可以免受潜在的后续故障的影响。因此，微网$h\in M$可由图$g_h=(N_{\mathrm{M},h},L_{\mathrm{M},h})$表示，$g_h\in G$且对于∀微网$h_1\in M$，∀微网$h_2\in M$，$h_1\neq h_2$，$g_{h_1}\cup g_{h_2}=\varnothing$。需要注意，这里的微网$h$泛指由微网$h$供电的配电系统分区。其中，$N_{\mathrm{M},h}$和$L_{\mathrm{M},h}$分别为微网$h$的节点集合和线路集合。

本文从DSO的视角来考虑DCLR问题，将微网视为一个整体，微网与配电系统之间的交互体现为公共连接点(point of common coupling，PCC)的功率传输[11,29]。时刻t微网h输出的有功功率和无功功率$p_{t,h}$和$q_{t,h}$需要满足微网容量约束和爬坡速率约束：

$$p_{t,h,\min}\leqslant p_{t,h}\leqslant p_{t,h,\max}\quad t\in\Gamma,h\in M \quad (7)$$
$$q_{t,h,\min}\leqslant q_{t,h}\leqslant q_{t,h,\max}\quad t\in\Gamma,h\in M \quad (8)$$
$$\begin{cases}p_{t,h,\min}=\max(p_{h,\min},p_{t-1,h}-\delta_{h,p}p_{h,\max})\\p_{t,h,\max}=\min(p_{h,\max},p_{t-1,h}+\delta_{h,p}p_{h,\max})\\q_{t,h,\min}=\max(q_{h,\min},q_{t-1,h}-\delta_{h,q}q_{h,\max})\\q_{t,h,\max}=\min(q_{h,\max},q_{t-1,h}+\delta_{h,q}q_{h,\max})\end{cases} \quad (9)$$

式中：$p_{t,h,\max}$和$p_{t,h,\min}$分别为时刻t微网h的有功功率上限和下限；$q_{t,h,\max}$和$q_{t,h,\min}$分别为时刻t微网h的无功功率上限和下限；$p_{h,\max}$和$p_{h,\min}$分别为微网h的有功功率、无功功率上限和下限；$q_{h,\max}$和$q_{h,\min}$分别为微网h的无功功率上限和下限；$\delta_{h,p}$和$\delta_{h,q}$分别为微网h的有功功率和无功功率爬坡速率，$\delta_{h,p}\in(0,1]$，$\delta_{h,q}\in(0,1]$。

3)用户满意度约束

用户满意度约束常见于电力系统可靠性研究[30-31]。从用户断电次数、用户断电电量等方面描述用户对于电力公司所提供电能服务的满意程度。本文在DCLR问题中考虑用户满意度对恢复决策的影响，假设在事后恢复期间，当2处负荷的重要程度(即加权负荷大小)相近时，DSO不会通过中断一处负荷的方式来恢复另一处负荷，即DSO会尽可能在维持已恢复负荷的基础上再考虑恢复其他负荷。假设用户满意度约束满足马尔可夫性质，即若给定当前状态，则未来状态与过去状态之间相互独立，该性质表述为：

$$P(s_{t+1}|s_t)=P(s_{t+1}|s_1,s_2,\cdots,s_t) \quad (10)$$

式中：$P(s_{t+1}|s_t)$为已知状态s_t条件下达到状态s_{t+1}的概率分布函数。

根据这一假设，且考虑到负荷重要程度对于恢复决策的影响已在恢复目标式(3)中体现，因此，DCLR中的用户满意度约束可通过对断开当前处于闭合状态的开关这一动作进行惩罚的方式来体现。

1.3 动态关键负荷恢复的MDP模型

针对DCLR数学模型，通过定义智能体、交互环境、控制动作、系统状态和奖励函数等要素来构建DCLR的MDP模型。

1)智能体和交互环境

DCLR问题中的智能体将承担DSO的角色，从配电系统的全局角度制定负荷恢复策略。与智能体交互的环境为具备动作执行、状态分析和奖励反馈等功能的配电系统模拟器。

2)控制动作

智能体在时刻t所采取的控制动作表述为：

$$a_t=\{o_{t,w}|t\in\Gamma,w\in W\} \quad (11)$$

式中：$o_{t,w}$为二值状态变量，表示时刻t针对开关w的操作，$o_{t,w}=0$表示闭合操作，$o_{t,w}=1$表示断开操作。由于动作空间A由有限的二值状态变量组成，因此，A具有“离散”的特点。

3）系统状态

在时刻t，系统状态由3类与DCLR决策主要相关的参数构成，即配电系统负荷参数（基于配电系统运行约束选取）、微网有功功率和无功功率上下限约束（基于微网运行约束选取）以及开关状态参数（基于用户满意度约束选取）。系统状态表述为：

$$s_t=\{p_{t,i,\phi},q_{t,i,\phi},p_{t,h,\min},p_{t,h,\max},q_{t,h,\min},q_{t,h,\max},o'_{t,w}|$$
$$t\in\Gamma,i\in N,\phi\in\{\mathrm{a},\mathrm{b},\mathrm{c}\},h\in M,w\in W\} \quad (12)$$

式中：$o'_{t,w}$为二值状态变量，表示开关w在时刻t的状态，$o'_{t,w}=0$表示闭合状态，$o'_{t,w}=1$表示断开状态。

由于状态空间S中包含功率类的连续变量，因此，S具有“连续”的特点。此外，s_t仅由部分可观参数构成，而其他参数（例如节点电压、线路电流）对决策的影响，可以假设智能体能够在与环境的交互过程中主动学习得到。

4）奖励函数

奖励函数直接影响智能体的决策行为，因此，奖励函数的设计需要综合考虑恢复目标和约束条件对于决策的影响。本文将约束条件分为2类：硬约束和软约束，如表1所示。恢复期间，若智能体决策违反软约束，则智能体将受到惩罚（即负值奖励），而负荷恢复可以继续进行；若违反硬约束，则智能体将受到严重惩罚，且负荷恢复失败。

表1　约束条件的类别
Table 1　Categories of constraint violations

约束条件	硬约束	软约束
功率平衡约束	是	
节点电压约束		是
线路电流约束	是	
辐射状网络拓扑约束	是	
微网容量约束	是	
微网爬坡速率约束	是	
用户满意度约束		是

奖励函数表述为：

$$r_t=-I_{0,t}c_0+(1-I_{0,t})(c_1\alpha_{1,t}-c_2\alpha_{2,t}-c_3\alpha_{3,t}) \quad (13)$$

$$\alpha_{1,t}=\sum_{i\in N,\phi\in\{\mathrm{a},\mathrm{b},\mathrm{c}\}}\omega_i\tilde{p}_{t,i,\phi} \quad (14)$$

$$\alpha_{2,t}=\sum_{i\in N,\phi\in\{\mathrm{a},\mathrm{b},\mathrm{c}\}}v_{t,i,\phi}(\max(0,V_{t,i,\phi}-V^{\max})+\max(0,V^{\min}-V_{t,i,\phi})) \quad (15)$$

$$\alpha_{3,t}=\sum_{w\in W}\max(0,o_{t,w}-o'_{t,w}) \quad (16)$$

式中：$I_{0,t}$为状态变量，表示时刻t的动作是否违反硬约束，$I_{0,t}=1$表示是，$I_{0,t}=0$表示否；$\alpha_{1,t}$为时刻t恢复的负荷总量；$\alpha_{2,t}$为表示节点电压约束违反情况的变量；$\alpha_{3,t}$为表示用户满意度约束违反情况的变量；c_0、c_1、c_2、c_3为系数，均为正系数，其中c_0远大于其他系数。

在式（13）中，$-c_0$表示违反硬约束的惩罚；$c_1\alpha_{1,t}$表示恢复负荷的奖励；$-c_2\alpha_{2,t}$表示违反节点电压约束的惩罚；$-c_3\alpha_{3,t}$表示违反用户满意度约束的惩罚。本文取$c_0=100$，$c_1=0.004\ \mathrm{kW}^{-1}$，$c_2=0.1$，$c_3=0.1$。在这一系数设置下，对于用户满意度约束而言，2处负荷的重要程度相近可以理解为2处负荷的有功功率在加权后的差距在25 kW以内（25 kW约为节点13与节点37系统中最小的单相负荷）。

1.4　基于OpenDSS的AEI实现

与智能体交互的DCLR模拟环境具备2个基本功能，即三相不平衡潮流计算和拓扑分析。其中，拓扑分析用于校验辐射状网络拓扑约束；三相不平衡潮流计算用于校验其他约束。s_t和r_t中的参数通过读取潮流计算结果的方式得到。

本文基于Python和OpenDSS构建确定性的DCLR模拟环境，进一步形成AEI。首先，在OpenDSS中搭建测试系统的三相不平衡潮流计算模型。其中，将微网建模为恒压源并连接至配电系统相应节点。此外，将配电系统的网络拓扑以图数据的格式进行存储，基于此，针对智能体做出的决策，利用深度优先搜索算法[32]校验拓扑约束。然后，通过OpenDSS组件接口，实现基于Python的DCLR模拟。在拓扑分析和潮流计算完成后，基于式（12）—式（16）形成系统状态和奖励并反馈给智能体。

DCLR的智能体-环境交互框架如图2所示。训练环节得到的智能体可以保存并用于应用环节，这一“离线训练-在线应用”的模式可以充分发挥神经网络输入输出映射的速度优势，从而有效提升应用环节的计算效率。

2　算法实现

强化学习是求解MDP的一般性框架。在所建立DCLR的MDP模型中，最优负荷恢复策略可表达为价值最高的控制动作。针对DCLR问题“状态空间连续-动作空间离散”的特点，采用DQN算法，通过值迭代的方式来搜索价值最高的控制动作。

DQN作为一种无模型的深度强化学习算法，在搜索最优策略的过程中无需额外的先验知识。

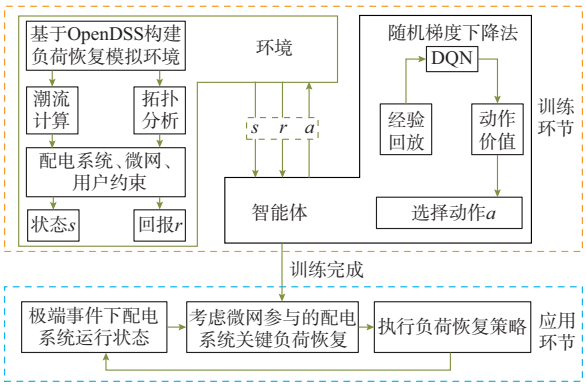

图2　负荷恢复问题的智能体-环境交互框架

Fig. 2　Agent-environment interaction framework of load restoration problem

2.1　DQN原理

最优控制策略可基于最优价值函数得到，最优动作价值函数$Q_*(s,a)$定义为：

$$Q_*(s,a)=\max_{\pi} E_{\tau\sim\pi}(R_{T\tau}|s_\tau=s,a_\tau=a) \tag{17}$$

最优动作价值函数遵循贝尔曼方程，若下一时刻状态s'中所有可能动作a'的最优价值$Q_*(s',a')$已知，则最优控制策略为选择动作a'使得期望动作价值$r+\gamma Q_*(s',a')$最大：

$$Q_*(s,a)=E_{s'}(r+\gamma\max_{a'}Q_*(s',a')|s_t=s,a_t=a) \tag{18}$$

进一步通过迭代更新的方式来估计第$\mu+1$次迭代时的动作价值函数$Q_{\mu+1}(s,a)$，如式(19)所示。

$$Q_{\mu+1}(s,a)=E_{s'}(r+\gamma\max_{a'}Q_*(s',a')|s_t=s,a_t=a) \tag{19}$$

根据贝尔曼最优性[17]，上述迭代过程最终将收敛至最优价值函数，即$Q_\mu(s,a)\to Q_*(s,a),i\to\infty$。由于难以单独估计每个状态-动作对的动作价值，因此，一般使用函数逼近的方法来估计动作价值函数。Q网络指具有权值参数θ的神经网络函数逼近器，可表示为$Q(s,a;\theta)\approx Q_*(s,a)$。

训练过程中通过迭代调整Q网络参数以逐步减小动作价值函数和目标价值函数之间的差距。最优目标动作价值$r+\gamma\max_{a'}Q_*(s',a')$由近似目标动作价值$y=r+\gamma\max_{a'}Q_*(s',a';\hat{\theta}_\mu)$表示，其中$\hat{\theta}_\mu$为前一次迭代中的Q网络参数。每次迭代的损失函数$F(\theta_\mu)$以及损失函数对于参数$\theta_\mu$的偏导表示为：

$$F(\theta_\mu)=E_{s,a,r,s'}((y-Q(s,a;\theta_\mu))^2) \tag{20}$$

$$\nabla_{\theta_\mu}F(\theta_\mu)=E_{s,a,r,s'}((y-Q(s,a;\theta_\mu))\nabla_{\theta_\mu}Q(s,a;\theta_\mu)) \tag{21}$$

针对损失函数$F(\theta_\mu)$，可基于随机梯度下降法等优化算法对参数θ进行迭代更新。

Q网络采用多层感知机模型，输入为系统状态，输出为各动作的动作价值，输出中最高动作价值对应的动作即为最优负荷恢复策略。

2.2　Q网络训练算法

DQN训练算法步骤如下：

步骤1：初始化经验回放池、动作价值网络Q以及目标动作价值网络$\hat{Q}$，episode取1。

步骤2：初始化环境状态s_t，$t=0$。

步骤3：基于ε-贪婪策略选择动作a_t。

步骤4：在DCLR环境中执行a_t，环境反馈奖励r_t和下一状态s_{t+1}，将经验(s_t,a_t,r_t,s_{t+1})存入经验回放池。

步骤5：从经验回放池中随机抽取小批量样本η_t。

步骤6：若$r_t=-c_0$，则近似目标动作价值$y_t=r_t$；否则$y_t=r_t+\gamma\max_{a'}Q_*(s_{t+1},a';\hat{\theta})$。

步骤7：针对$(y_t-Q(s_t,a_t;\theta))^2$，使用随机梯度下降法更新网络Q参数；每经过$C$次更新，将网络$\hat{Q}$替换为网络Q。

步骤8：若$t=T$或$r_t=-c_0$，则转至步骤9；否则，$t=t+1$，转至步骤3。

步骤9：若达到指定训练周期或是episode取到终止值，则转至步骤10；否则，episode自动加1，并转至步骤2。

步骤10：保存网络Q，结束训练。

图3为DQN训练算法原理图，其中，ε-贪婪策略表示智能体在做决策时，会以很小的概率ε随机选择一个动作，以1－ε的概率选择动作价值最大的动作，其中ε表示智能体的探索率。本文所使用的ε-贪婪模型表示如下：

$$\varepsilon=\varepsilon_\infty+(\varepsilon_0-\varepsilon_\infty)e^{-\frac{t_\varepsilon}{c_\varepsilon}} \tag{22}$$

式中：ε_0和ε_∞分别为初始和最终的探索率；t_ε为状态转移步数；c_ε为探索率衰减系数。

为保证算法的收敛性，DQN使用了经验回放和目标网络2个技巧，相关内容可参考文献[24-25]。

2.3　评价指标

智能体的收敛性指标和决策能力评价指标分别用于定量评价智能体在训练环节和应用环节的表现。收敛性指标包括：平均片段得分(average score

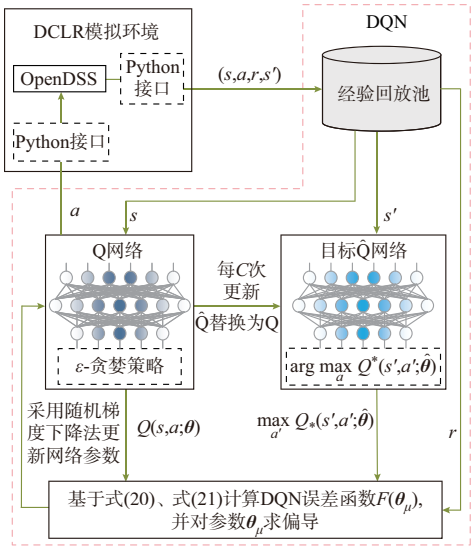

图3　DQN训练算法原理图

Fig. 3　Schematic diagram of DQN training algorithm

per episode，ASPE）[28]、平均动作价值（average action value，AAV）[28]和平均批次损失（average loss per batch，ALPB）。决策能力指标包括：平均成功率（average success rate，ASR）和平均最优性差距（average optimality gap，AOG）。

ASPE表示一个训练周期（epoch）内每个片段的平均折扣回报；AAV表示在一个训练周期内所选动作的平均价值；ALPB表示在一个训练周期内每个训练批次的平均训练损失。计算公式分别为：

$$\beta_{\mathrm{ASPE}}=\frac{1}{N_{\mathrm{e}}}\sum_{j=1}^{N_{\mathrm{e}}}\sum_{t=0}^{T_j-1}\gamma^t r_{j,t} \tag{23}$$

$$\beta_{\mathrm{AAV}}=\frac{1}{\sum_{j=1}^{N_{\mathrm{e}}}T_j}\sum_{j=1}^{N_{\mathrm{e}}}\sum_{t=0}^{T_j-1}Q(s_{j,t},a_{j,t};\boldsymbol{\theta}_j) \tag{24}$$

$$\beta_{\mathrm{ALPB}}=\frac{1}{N_{\mathrm{u}}}\sum_{j=1}^{N_{\mathrm{u}}}F(\boldsymbol{\theta}_j) \tag{25}$$

式中：N_{e}为训练周期内的片段数目；T_j为片段j的长度；N_{u}为1个训练周期包含的参数更新次数；$s_{j,t}$、$a_{j,t}$、$r_{j,t}$、$\boldsymbol{\theta}_j$分别为片段j的状态、动作、立即奖励和参数。

针对应用环节，考虑到DRL具有黑箱问题，本文从动作的可行性和最优性2个方面来衡量智能体的决策能力。在决策能力指标中，ASR表示智能体采取可行动作的概率，ASR在[0,1]范围内取值，ASR越趋近于1，表明智能体越为可靠；AOG衡量基于DRL的决策解和基准解在恢复目标方面的差距，AOG越接近于0，表明所提方案与基准方案在最优性方面越接近。

对于ASR指标，调用Q网络生成多个具有长度上限的决策片段，长度上限用于避免生成过长的片段，因此，ASR定义为生成片段中可行动作的占比：

$$\beta_{\mathrm{ASR}}=\left(1-\frac{\sum_j R_j}{\sum_j T_j}\right)\times 100\% \tag{26}$$

式中：R_j为辨识片段j是否以不可行动作结束的变量，$R_j=1$表示是，$R_j=0$表示否。

对于AOG指标，调用Q网络生成多个固定长度的决策片段，片段中均不含不可行动作。对于每个片段，基于DRL的决策方案和基准方案之间的最优性差距β_{gap}表示为：

$$\beta_{\mathrm{gap}}=1-\frac{\sum_{t\in\Gamma,i\in N,\phi\in\{\mathrm{a,b,c}\}}\omega_i\tilde{p}_{t,i,\phi}}{\sum_{t\in\Gamma,i\in N,\phi\in\{\mathrm{a,b,c}\}}\omega_i\tilde{p}_{t,i,\phi,\mathrm{B}}} \tag{27}$$

式中：$\tilde{p}_{t,i,\phi,\mathrm{B}}$为基准方案中实际恢复的负荷。AOG指标则为生成片段的平均最优性差距。

3　算例分析

将所提方法应用于2个改进的IEEE测试系统，即IEEE 13节点馈线系统（本文简称13节点系统）[33]和IEEE 37节点馈线系统（本文简称37节点系统）[34]，以验证其有效性。所有计算均在搭载Intel Core i7-8700 3.20 GHz CPU，8 GB RAM的计算机上进行。所提方案与基准方案2的环境配置为Python 3.6.12、PyTorch 1.8.0和OpenDSS 9.1.0.1；基准方案1的环境配置为MATLAB和Gurobi 9.1.0。本项工作未使用GPU加速计算和分布式训练。

3.1　测试系统

13节点系统中设置了2个微网，分别接在节点633和692；同时设置了4个分段开关用于执行负荷恢复决策，分别安装在线路632-645、670-671、671-684和671-692。37节点系统中设置了4个微网和6个用于执行负荷恢复决策的分段开关。13节点与37节点系统的三相拓扑见附录A。此外，假设下一时刻的负荷在当前负荷的基础上加上±30%的随机波动，微网的爬坡速率为100%。测试系统的具体数据，例如负荷参数、微网参数、状态空间、动作空间的设置等分别见附录B和C。

3.2　训练效果

训练环节首先需要确定各测试系统所对应的Q网络的架构。经过尝试，最终确定Q网络由4个全连接线性层组成，每个线性层后均接有修正线性单

元(rectified linear unit,ReLU)。输入层的神经元数目等于状态s中的参数数目(13节点系统的输入层神经元数量为42;37节点系统的输入层神经元数量为82)。两个隐藏层均包括512个神经元。输出层的神经元数目等于动作空间A中动作的数目(13节点系统的输出层神经元数目为16;37节点系统输出层神经元数目为30)。然后,基于DQN算法训练Q网络,训练环节中的超参数取值见附录D。此外,本文中所使用的随机梯度下降算法为Adam[35]。

基于附录B和C中测试系统参数设置和附录D中超参数设置,针对折扣因子γ的取值,设置如表2所示的3个测试场景,用于分析DCLR问题中智能体"近视"程度对于训练和应用效果的影响。

表2　测试场景设置
Table 2　Setup of test scenarios

场景	折扣因子γ	设置内容
1	0.50	智能体考虑未来约3步决策的奖励
2	0.90	智能体考虑未来约20步决策的奖励
3	0.99	智能体考虑未来约200步决策的奖励

图4展示了训练过程中收敛性指标的变化情况。其中,ASPE的变化趋势可基于片段长度和立即奖励大小来分析。

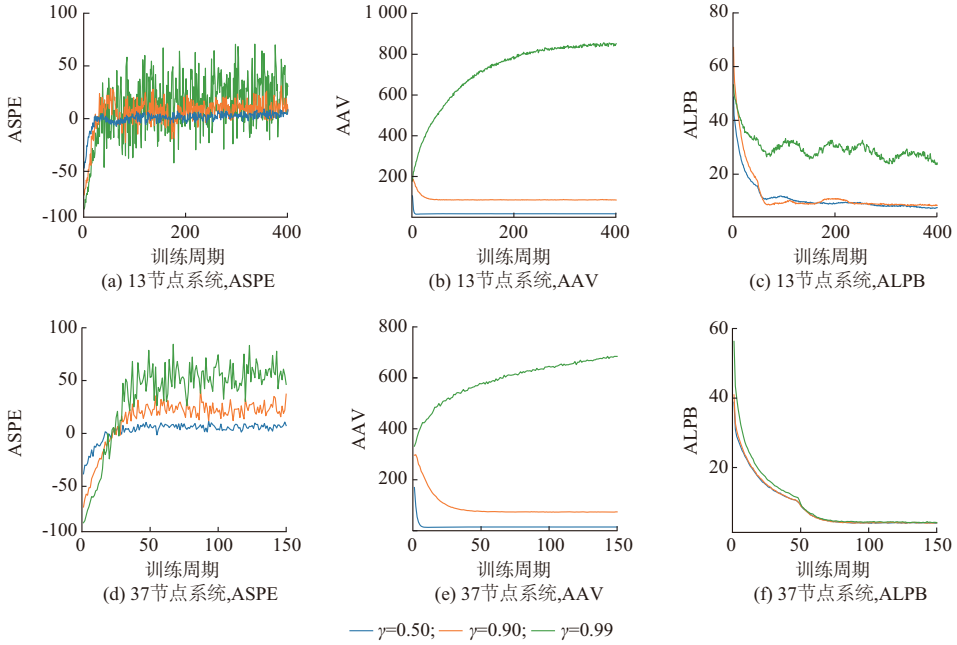

图4　ASPE、AAV、ALPB训练曲线
Fig. 4　Training curves of ASPE, AAV, and ALPB

ASPE前期收敛较快表示智能体经过较少训练周期就能学习到可行控制策略。此后,智能体会采取一些回报更高但失败风险也更大的动作以提升立即奖励,这可能使得恢复片段提前终止,并伴随负值奖励作为惩罚。因此,后期当风险与收益没有得到很好的权衡时,ASPE曲线会出现明显波动。为避免训练发散,可在优化算法中适当选取较小的学习率(见附录D表D1)。此外,当γ取值越小时,由于智能体越为"近视",因此,片段长度会更短,ASPE的收敛值也更小,如图4所示。需要说明的是,DQN算法对回放池中负样本(即违反硬约束的经验)的比例比较敏感,而ASPE较小意味着负样本比例会相应地变大,所以本文中的折扣因子取值不宜过小。

对于AAV,通过监测训练过程中Q网络的输出发现,所有动作的价值随时间变化的趋势分为2类:一类会很快收敛至0,这对应于不可行动作(即违反硬约束的动作);另一类对应于可行动作,其价值逐渐收敛至期望折扣回报,且所有可行动作具有相似的变化趋势。这解释了AAV变化趋势的平稳特征。对于场景1和2,γ取值较小,使得期望折扣回报较小,且低于Q网络初始值,AAV呈现下降趋势;而场景3中期望折扣回报高于Q网络初始值,AAV呈现上升趋势。此外,对于ALPB,该指标与Q网络损失直接相关,可较为直观地体现智能体的训练难度。

此外,虽然13节点系统的输入与输出参数的数目少于37节点系统,但由于其动作空间包含了所有

可能的开关动作组合（见附录B），而37节点系统的动作空间中已提前删去了违反拓扑约束的34个开关动作组合（见附录C），使得13节点系统的动作空间中存在较多不可行动作（在基础负荷下，13节点系统动作空间中不可行动作占比为1/2；37节点系统为2/5）；另一方面，相比于37节点系统，13节点系统中的关键负荷总量与微网容量较为接近（见附录B和C）。因此，在类似的训练环节下，37节点系统的“学习难度”反而小于13节点系统，其训练效果也更优，如图4所示。

训练结果表明13节点和37节点测试系统的智能体均可以取得较好的收敛结果，且收敛性指标在训练初期的收敛速度较快。针对同一系统的不同测试场景，γ取值越小，则训练环节的收敛效果越好。这与1.2节的分析一致，即γ越小表示智能体在决策时向前考虑的步数越少、训练难度越小。此外，ASPE和ALPB的收敛特性较为相似，AAV的变化趋势更为平滑。在计算效率方面，对于13节点系统，训练一个训练周期数目为500的智能体需要5 h左右；对于37节点系统，训练一个训练周期数目为150的智能体需要2 h左右。

3.3　应用效果

在应用环节中，将系统状态输入训练好的Q网络，并选择输出中最高动作价值对应的动作作为每次负荷恢复决策的最优解。本文将所提方案与3个基准方案进行对比来分析其在决策最优性和计算效率方面的表现。基准方案1为文献[29]所提的基于模型的求解方法，其中DCLR问题被建模为混合整数线性规划（mixed-integer linear programming，MILP）问题。基准方案2和3将DCLR问题视为各时间段的单阶段静态关键负荷恢复问题的直接组合。每个单阶段静态关键负荷恢复问题通过在OpenDSS中遍历动作空间，并将其中恢复关键（加权）负荷最多的动作视为最优解。所提方案与基准方案之间的主要区别如表3所示。决策能力指标结果如表4所示，可以发现γ取值较大时ASR和AOG指标更优，这意味着DCLR问题中γ较大时智能体的决策能力更优。

对于ASR，场景3下13节点系统与37节点系统的智能体分别达到99.99%和99.98%。实际应用中为保证系统的安全运行，需要对基于DRL的决策解进行安全校验。本文设定如下校验机制：当最优解违反硬约束时，选择次优解（即Q网络输出中第二高动作价值所对应的动作），并对其进行校验。该校验过程可以重复直到校验通过为止。实验发现，在仅设置一轮安全校验的情况下（即只对最优解进行校验），场景3下2个测试系统的ASR指标普遍达到100%。

表3　恢复方案之间的区别
Table 3　Differences between restoration methods

方案	数学模型	潮流模型	约束条件
所提方案	MDP	三相不平衡潮流	考虑用户满意度
基准方案1	MILP	线性配电网潮流	考虑用户满意度
基准方案2	枚举	三相不平衡潮流	考虑用户满意度
基准方案3	枚举	三相不平衡潮流	忽略用户满意度

表4　应用环节的指标结果
Table 4　Index results of application segment

系统	场景	ASR/%	AOG1/%	AOG2/%	AOG3/%
13节点系统	1	99.52	3.13	2.850 0	2.930 0
	2	99.62	2.93	2.630 0	2.690 0
	3	99.99	2.21	1.910 0	1.960 0
37节点系统	1	99.95	1.12	0.010 0	0.020 0
	2	99.98	1.11	0.010 0	0.020 0
	3	99.98	1.08	0.000 1	0.000 2

注：ASR中，片段数目为10，片段长度上限为100；AOG中，片段数目为10，片段长度为100，AOG1、AOG2、AOG3分别为所提方案针对基准方案1、2、3的AOG指标，AOG指标对应的恢复负荷量见附录E。

对于AOG，所提方案与基准方案的结果较为接近。实验发现，相同恢复策略下，基准方案2的潮流计算结果中，部分节点的负荷略低于基准方案1中所得到的结果，即潮流模型的差异（见表3）使得AOG1大于AOG2。为进一步分析误差来源，针对13节点系统和场景3，选择了一组分别由所提方案和基准方案得到的决策片段，如附录E图E1所示，并从每次状态转换的立即奖励和恢复负荷2个方面分析最优性偏差的来源。对比所提方案和基准方案2，可以发现最优性偏差主要来源于智能体的保守性决策，例如附录E图E2中的第84次和第87次状态转换；对比所提方案和基准方案3，可以发现考虑到用户满意度约束，智能体会做出一些回报较高但恢复负荷较小的决策，例如附录E图E2中的第43次和第55次状态转换（附录E中讨论了第42次和第43次状态转换对应的负荷恢复过程）。前者反映了智能体的风险回避特性，需要通过改进训练算法等方式来提升智能体的决策能力；而后者是因为所提方案和基准方案3在决策目标上具有差异，这体现了智能体在处理复杂问题时的有效性。

应用环节的计算效率对比如表5和表6所示，其中计算效率定义为平均单次决策耗时。对于所提方案，借助于Q网络的输入输出映射关系，其在应用

环节极具效率优势，且计算效率随系统规模的变化而降低。由于动作空间离散且动作数目较少，因此，基准方案2和3（基于枚举）的计算效率与基准方案1相当，均在秒级左右。

表5 应用环节的计算效率
Table 5 Computational efficiency of application segment

系统	计算时间/s			
	所提方案	基准方案1	基准方案2	基准方案3
13节点系统	1.49×10^{-4}	0.756 6	0.163 9	0.163 9
37节点系统	1.50×10^{-4}	1.211 5	0.481 9	0.481 9

表6 应用环节的计算效率比值
Table 6 Computational efficiency ratio of application segment

系统	效率比值1	效率比值2	效率比值3
13节点系统	5 078	1 100	1 100
37节点系统	8 077	3 213	3 213

注：效率比值1、2、3分别为基准方案1、2、3的计算效率与所提方案计算效率的比值。

离线训练得到的Q网络仍然可以根据应用需求来滚动更新其网络参数，以进一步保障基于DRL的决策方案在实际应用中的适用性。

4 结语

本文提出了一种基于DRL的面向弹性提升的含微网配电系统动态关键负荷恢复方法，可以在极端事件导致主网供电能力不足时为配电系统中的关键负荷恢复提供解决方案。基于智能体-环境交互过程，所提方法支持端到端的决策方式。算例结果表明，基于DRL的方法可以成功地学习负荷恢复策略，且无需额外的先验知识。在实际应用中，通过引入安全校验，可以有效保证决策的成功率和系统的安全运行。对于决策的最优性，尽管智能体还存在一定的决策保守性，但整体而言，所提方法具有较好的表现。此外，所提方法在应用环节的计算效率方面具有明显优势。

同时，本文工作也存在一定局限性，进一步的研究将处理更为复杂的恢复场景和恢复动作，并尝试量化恢复期间的不确定性因素对恢复结果的影响。此外，还将针对DRL算法、高性能计算，以及面向DRL应用的标准化负荷恢复模拟环境及其开源等问题展开研究。

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] BENNETT J A, TREVISAN C N, DECAROLIS J F, et al. Extending energy system modelling to include extreme weather risks and application to hurricane events in Puerto Rico [J]. Nature Energy, 2021, 6(3): 240-249.

[2] 别朝红，林雁翎，邱爱慈. 弹性电网及其恢复力的基本概念与研究展望[J]. 电力系统自动化，2015，39(22)：1-9.
BIE Zhaohong, LIN Yanling, QIU Aici. Concept and research prospects of power system resilience [J]. Automation of Electric Power Systems, 2015, 39(22): 1-9.

[3] 别朝红，林超凡，李更丰，等. 能源转型下弹性电力系统的发展与展望[J]. 中国电机工程学报，2020，40(9)：2735-2745.
BIE Zhaohong, LIN Chaofan, LI Gengfeng, et al. Development and prospect of resilient power system in the context of energy transition [J]. Proceedings of the CSEE, 2020, 40(9): 2735-2745.

[4] 王守相，刘琪，赵倩宇，等. 配电网弹性内涵分析与研究展望[J]. 电力系统自动化，2021，45(9)：1-9.
WANG Shouxiang, LIU Qi, ZHAO Qianyu, et al. Connotation analysis and prospect of distribution network elasticity [J]. Automation of Electric Power Systems, 2021, 45(9): 1-9.

[5] PANTELI M, MANCARELLA P. Influence of extreme weather and climate change on the resilience of power systems: impacts and possible mitigation strategies [J]. Electric Power Systems Research, 2015, 127: 259-270.

[6] 葛少云，张成昊，刘洪，等. 考虑微能源网支撑作用的配电网弹性提升策略[J]. 电网技术，2019，43(7)：2306-2317.
GE Shaoyun, ZHANG Chenghao, LIU Hong, et al. Resilience enhancement strategy for distribution network considering supporting role of micro energy grid [J]. Power System Technology, 2019, 43(7): 2306-2317.

[7] 周晓敏，葛少云，李腾，等. 极端天气条件下的配电网韧性分析方法及提升措施研究[J]. 中国电机工程学报，2018，38(2)：505-513.
ZHOU Xiaomin, GE Shaoyun, LI Teng, et al. Assessing and boosting resilience of distribution system under extreme weather [J]. Proceedings of the CSEE, 2018, 38(2): 505-513.

[8] NAZEMI M, MOEINI-AGHTAIE M, FOTUHI-FIRUZABAD M, et al. Energy storage planning for enhanced resilience of power distribution networks against earthquakes [J]. IEEE Transactions on Sustainable Energy, 2020, 11(2): 795-806.

[9] WANG J, ZUO W D, RHODE-BARBARIGOS L, et al. Literature review on modeling and simulation of energy infrastructures from a resilience perspective [J]. Reliability Engineering & System Safety, 2019, 183: 360-373.

[10] MISHRA D K, GHADI M J, AZIZIVAHED A, et al. A review on resilience studies in active distribution systems [J/OL]. Renewable and Sustainable Energy Reviews, 2021, 135 [2021-08-11]. https://www.sciencedirect.com/science/article/abs/pii/S1364032120304913.

[11] CHEN C, WANG J H, QIU F, et al. Resilient distribution system by microgrids formation after natural disasters [J]. IEEE Transactions on Smart Grid, 2016, 7(2): 958-966.

[12] 许寅，王颖，和敬涵，等. 多源协同的配电网多时段负荷恢复优

化决策方法[J].电力系统自动化，2020，44(2)：123-131.
XU Yin, WANG Ying, HE Jinghan, et al. Optimal decision-making method for multi-period load restoration in distribution network with coordination of multiple sources[J]. Automation of Electric Power Systems, 2020, 44(2): 123-131.

[13] XU Y, LIU C C, SCHNEIDER K P, et al. Microgrids for service restoration to critical load in a resilient distribution system[J]. IEEE Transactions on Smart Grid, 2018, 9(1): 426-437.

[14] 卞艺衡，别朝红.面向弹性提升的智能配电网远动开关优化配置模型[J].电力系统自动化，2021，45(3)：33-39.
BIAN Yiheng, BIE Zhaohong. Resilience-enhanced optimal placement model of remote-controlled switch for smart distribution network [J]. Automation of Electric Power Systems, 2021, 45(3): 33-39.

[15] SHI Q X, LI F X, OLAMA M, et al. Network reconfiguration and distributed energy resource scheduling for improved distribution system resilience [J/OL]. International Journal of Electrical Power & Energy Systems, 2021, 124 [2021-05-12]. https://doi.org/10.1016/j.ijepes.2020.106355.

[16] SHI Q X, LI F X, OLAMA M, et al. Post-extreme-event restoration using linear topological constraints and DER scheduling to enhance distribution system resilience [J/OL]. International Journal of Electrical Power & Energy Systems, 2021, 131[2021-05-12]. https:// doi.org/ 10.1016/ j.ijepes.2021.107029.

[17] GILANI M A, KAZEMI A, GHASEMI M. Distribution system resilience enhancement by microgrid formation considering distributed energy resources [J/OL]. Energy, 2020, 191[2021-05-12]. https:// doi.org/ 10.1016/ j.energy.2019.116442.

[18] 刘礼邦，武传涛，随权，等.计及可控负荷参与的主动配电网动态恢复供电策略[J].电力系统保护与控制，2020，48(9)：27-35.
LIU Libang, WU Chuantao, SUI Quan, et al. Power supply strategy for active distribution network dynamic recovery with controllable load participation[J]. Power System Protection and Control, 2020, 48(9): 27-35.

[19] FARZIN H, FOTUHI-FIRUZABAD M, MOEINI-AGHTAIE M. Enhancing power system resilience through hierarchical outage management in multi-microgrids[J]. IEEE Transactions on Smart Grid, 2016, 7(6): 2869-2879.

[20] CAI S, XIE Y Y, WU Q W, et al. Robust MPC-based microgrid scheduling for resilience enhancement of distribution system [J/OL]. International Journal of Electrical Power & Energy Systems, 2020, 121[2021-05-12]. https://doi.org/10.1016/j.ijepes.2020.106068.

[21] 赵俊华，董朝阳，文福拴，等.面向能源系统的数据科学：理论、技术与展望[J].电力系统自动化，2017，41(4)：1-11.
ZHAO Junhua, DONG Zhaoyang, WEN Fushuan, et al. Data science for energy systems: theory, techniques and prospect[J]. Automation of Electric Power Systems, 2017, 41(4): 1-11.

[22] BELLMAN R. Dynamic programming[J]. Science, 1966, 153(3731): 34-37.

[23] Electric Power Research Institute. Open distribution system simulator (OpenDSS)[EB/OL]. [2021-03-12]. https://www.epri.com/pages/sa/opendss.

[24] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning[R/OL]. [2021-03-12]. https://arxiv.org/abs/1312.5602.

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

[26] SCHNEIDER K, TUFFNER F, ELIZONDO M, et al. Evaluating the feasibility to use microgrids as a resiliency resource[J]. IEEE Transactions on Smart Grid, 2016, 8(2): 687-696.

[27] CALDOGNETTO T, TENTI P. Microgrids operation based on master-slave cooperative control [J]. IEEE Journal of Emerging and Selected Topics in Power Electronics, 2014, 2(4): 1081-1088.

[28] PARHIZI S, LOTFI H, KHODAEI A, et al. State of the art in research on microgrids: a review[J]. IEEE Access, 2015, 3: 890-925.

[29] DING T, LIN Y L, LI G F, et al. A new model for resilient distribution systems by microgrids formation [J]. IEEE Transactions on Power Systems, 2017, 32(5): 4145-4147.

[30] SULLIVAN M J, NOLAND S B, VARDELL T, et al. Interruption costs, customer satisfaction and expectations for service reliability[J]. IEEE Transactions on Power Systems, 1996, 11(2): 989-995.

[31] LI G F, BIE Z H, XIE H P, et al. Customer satisfaction based reliability evaluation of active distribution networks[J]. Applied Energy, 2016, 162: 1571-1578.

[32] WEST D B. Introduction to graph theory[M]. Upper Saddle River, USA: Prentice Hall, 2001.

[33] IEEE PES AMPS DSAS Test Feeder Working Group. IEEE 13-bus feeder[EB/OL]. [2021-02-12]. http:// site.ieee.org/pestestfeeders/files/2017/08/feeder13.zip.

[34] IEEE PES AMPS DSAS Test Feeder Working Group. IEEE 37-bus feeder[EB/OL]. [2021-02-12]. http://site.ieee.org/pes-testfeeders/files/2017/08/feeder37.zip.

[35] KINGMA D P, BA J. Adam: a method for stochastic optimization[C]// 3rd International Conference on Learning Representations, May 7-9, 2015, San Diego, USA.

黄玉雄(1995—)，男，博士研究生，主要研究方向：电力系统可靠性、人工智能在电力系统的应用、综合能源系统。E-mail：hyx.xj@stu.xjtu.edu.cn

李更丰(1983—)，男，博士，副教授，主要研究方向：电力系统可靠性、综合能源系统与主动配电网技术。E-mail：gengfengli@xjtu.edu.cn

张理寅(1999—)，男，硕士研究生，主要研究方向：弹性电力系统、人工智能在电力系统的应用。E-mail：zhangliyin@stu.xjtu.edu.cn

别朝红(1970—)，女，通信作者，博士，教授，主要研究方向：电力系统规划及可靠性评估、新能源电力系统安全风险评估、弹性电力系统。E-mail：zhbie@mail.xjtu.edu.cn

（编辑　顾晓荣）

Deep Reinforcement Learning Method for Dynamic Load Restoration of Resilient Distribution Systems

HUANG Yuxiong, *LI Gengfeng*, *ZHANG Liyin*, *BIE Zhaohong*

(School of Electrical Engineering, Xi′an Jiaotong University, Xi′an 710049, China)

Abstract: In extreme events, it can effectively enhance the grid resilience by using surplus power of microgrids to serve the critical loads in distribution systems. Based on the deep reinforcement learning (DRL) technique, considering the participation of microgrids, a dynamic critical load restoration (DCLR) method of distribution systems is proposed to support the model-free manner to solve the complex problems, significantly improving the online computational efficiency. Firstly, the DCLR problem of distribution systems with microgrids is analyzed. On this basis, its Markov decision process (MDP) is formulated considering the complex operational constraints, including the distribution operation constraints, microgrid operation constraints, customer satisfaction degree, etc. Secondly, a DCLR simulation environment is built based on OpenDSS as the agent-environment interface for applying DRL algorithms. Furthermore, a deep Q-network algorithm is adopted to search for the optimal policies of the critical load restoration. The convergence and decision-making ability indices are defined to measure the performance of the agents in training and application processes, respectively. Based on the two modified IEEE test systems, the effectiveness of the proposed method is verified.

This work is supported by State Grid Corporation of China (No. SGJX0000KXJS1900322).

Key words: load restoration; deep learning; reinforcement learning; distribution system; microgrid; resilience