第46卷 第23期 2022年12月10日
DOI：10.7500/AEPS20210824002
电力系统自动化
Automation of Electric Power Systems
Vol. 46 No. 23 Dec. 10，2022

基于态势利导的需求响应自学习优化调度方法

明威宇，李　妍，程时杰，龙　禹，徐　菁，王少荣
（强电磁工程与新技术国家重点实验室，华中科技大学，湖北省武汉市 430074）

摘要：针对多随机场景下用户可选择需求响应(CCR)的场景组合激增问题，利用深度强化学习算法实现CCR群组的优选及其所包含节点的优化调度。首先，根据CCR优化调度的约束条件与目标函数，分析其数学模型及日调度周期的求解复杂度；然后，基于马尔可夫决策过程将CCR优化调度过程映射至态势感知元组，并基于竞争深度Q网络架构建立态势利导函数，通过多次态势推演，利用小批量梯度下降法对态势利导函数求导，不断反馈更新算法参数，实现决策优化；最后，基于IEEE 33节点算例，通过不同规模的随机样本数量，在随机运行方式下实现了待选CCR群组的优选，并制定相应的优化调度策略。

关键词：可选择需求响应；深度强化学习；竞争深度Q网络；马尔可夫决策过程；态势感知；态势利导

0　引言

随着中国电力市场化改革的快速推进[1]，用户可选择需求响应(consumer choice resource，CCR)基于自身意愿主动参与到电力市场各项业务中[2-3]。通过对CCR的调度，可以将负荷侧资源配合电网运行加以充分利用，从而减少网损[4]、提升设备使用寿命[5]、改善用户的用电体验[6]，在满足网侧精益化管理的同时实现用户侧降费提质的需求。但CCR受用户主观意愿和负荷动态物理特性等多因素影响[7-8]，其优化调度需要考虑多目标综合优化和系统运行的安全约束，协同众多变量优化求解，其优化问题为具有复杂动态约束的混合整数非线性规划模型，在配电网随机运行方式下求解时，存在场景组合激增的问题，求解的复杂度随求解时段数成指数增长，难以找到最优解[9-10]。

随着近年来数据驱动的机器学习方法的发展[11-12]，深度强化学习(deep reinforcement learning，DRL)在多个领域的序贯决策优化问题中得到了广泛应用[13-15]。已有不少学者利用DRL将电力系统随机优化决策问题映射至马尔可夫决策过程(Markov decision process，MDP)模型，以自学习方式予以求解。文献[16]对DRL应用于需求响应业务的可行性与方法进行了探讨，提出了基于DRL的需求响应业务开展架构。文献[17-18]关注到需求响应业务侧负荷的联合竞价及定价问题，利用基于DRL的深度确定性策略梯度方法[18]，基于MDP对负荷的联合竞价及定价问题进行建模，建立动态竞价响应函数，通过自学习历史数据优化终端用户用电行为。文献[19]利用改进深度确定性策略梯度算法计算楼宇级控制策略，建立调度中心-负荷聚集商-楼宇级控制单元-用户的调度架构，将电采暖动作、用户费用及调度成本等纳入MDP，从而基于DRL调度用电采暖参与需求响应。文献[20]依托演员-批评家结构的DRL算法，将工业设施中储能设备的电能状态、工业设备动作情况纳入MDP，利用DRL制定工业设施的最佳能源管理策略，实现需求响应侧业务优化管理。文献[21]基于DRL将用户不满意度、售电商经济收益纳入MDP，实现了激励型需求响应的补贴价格决策优化。文献[22]将电动汽车作为需求响应资源，将电动汽车充放电动作、电网功率波动值等情况纳入MDP，基于DRL实现了需求响应的优化决策。综上所述，DRL求解CCR优化问题的有效性已得到广泛关注。

本文基于态势利导的需求响应自学习优化调度方法，首先，分析以电压安全运行为约束条件，以供电公司经济补偿和停电次数最小为目标的CCR群组节点优化调度数学模型；然后，构建MDP模型的CCR群组节点态势感知元组和态势利导函数；进而，通过对历史负荷数据曲线的泛化处理，DRL算

收稿日期：2021-08-24；修回日期：2022-04-22。
上网日期：2022-10-18。
国家重点研发计划智能电网技术与装备重点专项资助项目(2017YFB0902800)。

法在ε-greedy策略和经验池机制下训练态势利导函数，以预测电网运行状态以及模拟用户行为，通过自趋优决策实现多组待选CCR群组的优选及其所包含节点的优化调度；最后，以IEEE 33节点为算例，对比分析竞争深度Q网络(dueling deep Q network，DDQN)结构和深度Q网络(deep Q network，DQN)结构的CCR群组优选求解策略，体现了DDQN结构DRL算法的优越性，对比DDQN结构下不同规模的样本数量的CCR群组优选求解策略，验证了所提方法适应多时间断面复杂场景的有效性。

1　需求响应优化调度的数学模型

在保证CCR群组节点响应后电压运行在安全范围的前提下，供电公司因CCR群组节点调度给予用户经济补偿将影响其售电利润，且用户侧停电次数不能过多，因此优化模型目标为电网经济补偿与停电次数最小。优化调度的目标函数如式(1)所示，其中第1项为供电公司经济补偿函数，第2项为停电次数惩罚函数，由于两者量纲不同，且数值存在数量级差距，故将其归一化处理。考虑到当CCR群组节点响应后，电网节点电压应运行在合理范围内，电压运行惩罚函数如式(2)所示。

$$f=\min\left(\frac{\sum_{j=1}^{n}\int_0^T\lambda_t a_t^{(j)}P_t^{(j)}\mathrm{d}t}{\sum_{j=1}^{n}\int_0^T\lambda_t P_t^{(j)}\mathrm{d}t}+\frac{\sum_{j=1}^{n}\int_0^T a_t^{(j)}\mathrm{d}t}{nT}\right) \tag{1}$$

$$\text{s.t.}\quad \min\left(\int_0^T\left[U_{t+1}^{(j)}-0.93U_{\mathrm{e}}\right]\left|a_{t+1}^{(j)}-a_t^{(j)}\right|\mathrm{d}t\right) \tag{2}$$

式中：T为日调度的一个周期；n为CCR群组节点数；λ_t为t时刻电价；$P_t^{(j)}$为t时刻第j号CCR群组节点的核定削减功率；$a_t^{(j)}$和$a_{t+1}^{(j)}$分别为第j号CCR群组节点在t和$t+1$时刻的响应状态；[·]为取整函数；$U_{t+1}^{(j)}$为第j号CCR群组节点在$t+1$时刻的电压标幺值；U_{e}为额定电压标幺值。

本文通过CCR群组节点的组合优化控制实现优化目标，优化变量为：

$$A=\{a_t^{(j)}\}\qquad j\in N_{\mathrm{CCR}} \tag{3}$$

式中：N_{CCR}为CCR群组节点集合。本文定义响应状态集合为{响应，未响应}。

在日调度周期T中，CCR群组节点(即功率可观测节点)有n个，在其响应后对c个节点电压进行观测，在每个时间断面的运行方式所满足的潮流约束见附录A，针对c个节点的电压，需要针对2^n个数据样本，在2^n个状态空间中选择一组优化状态。况且日调度周期T中如果有w个时间断面，考虑到相邻时间断面的停电次数和供电公司售电利润的优化目标，故在一个周期内，需针对2^{nw}个数据样本，在2^{nw}个状态空间中选择一组优化状态。因此，电网运行状态随机性会导致场景组合激增，求解的复杂度随求解时段数呈指数增长，优化模型难以找到最优解。

2　态势感知元组及态势利导函数

本章基于MDP建立自学习智能体态势感知元组(S,A,R)，其中S为态势感知获取的状态集，A为响应状态动作集，R为环境理解函数，基于态势感知元组构建态势利导函数，通过自趋优态势利导实现CCR群组的优选及其所包含节点的调度优化。

1)态势感知获取的状态集S

以配电网节点电压和CCR群组节点的响应功率为感知量，配电网状态和CCR群组中节点的状态构成状态集S，如式(4)所示。

$$S=\{U_{\mathrm{END},t}^{(i)},P_{\mathrm{CCR},t}^{(j)}\}\qquad i\in N_{\mathrm{END}},j\in N_{\mathrm{CCR}} \tag{4}$$

式中：N_{END}为可观测电压节点的集合；$U_{\mathrm{END},t}^{(i)}$为t时刻可观测节点i的电压；$P_{\mathrm{CCR},t}^{(j)}$为CCR群组中t时刻节点j的响应功率。

2)响应状态动作集A

响应状态即为式(3)所示优化变量。$a_t^{(j)}$取值为0(CCR群组节点响应)或1(CCR群组节点未响应)。

3)环境理解函数R

为实现CCR群组优化调度目标，建立的环境理解函数R包括供电公司售电利润函数、响应状态函数以及电压运行回报函数，如式(5)所示。

$$R_{t+1}=\sum_{i\in N_{\mathrm{END}}}r_{i,t+1}^{\mathrm{u}}+\sum_{j\in N_{\mathrm{CCR}}}r_{j,t+1}^{\mathrm{a}}+r_{t+1}^{\mathrm{DSO}} \tag{5}$$

式中：R_{t+1}为在$t+1$时刻的环境理解函数值，反映上一时刻响应状态的优劣。

对于电压运行回报函数$r_{i,t+1}^{\mathrm{u}}$，当CCR群组节点响应后，可观测节点电压位于合理范围内时，电压运行回报函数取正向激励(值)F^{u}；反之，取0。电压运行回报函数为：

$$r_{i,t+1}^{\mathrm{u}}=\begin{cases}F^{\mathrm{u}} & 0.93U_{\mathrm{e}}<U_{t+1}^{(i)}<1.07U_{\mathrm{e}}\\ 0 & \text{其他}\end{cases} \tag{6}$$

对于响应状态函数$r_{j,t+1}^{\mathrm{a}}$，若相邻时刻开关状态一致，则$r_{j,t+1}^{\mathrm{a}}$取0；反之，取F_j^{a}。响应状态函数为：

$$r_{j,t+1}^{\mathrm{a}}=\begin{cases}F_j^{\mathrm{a}} & a_{t+1}^{(j)}\neq a_t^{(j)},j\in N_{\mathrm{CCR}}\\ 0 & \text{其他}\end{cases} \tag{7}$$

F_j^{a}与节点期望停电次数有关，其定义如式(8)所示。

$$F_j^{\mathrm{a}}=\frac{k}{k_j} \tag{8}$$

式中：k_j为节点期望停电次数；k为固定正常数。

对于供电公司售电利润函数r_{t+1}^{DSO}，当负荷动作使得供电公司售电利润大于补偿时，则r_{t+1}^{DSO}取0，反之取负向激励（值）F^{p}。供电公司售电利润函数为：

$$r_{t+1}^{\mathrm{DSO}}=\begin{cases}F^{\mathrm{p}} & M_{t+1}^{\mathrm{pr}}-M_{t+1}^{\mathrm{co}}<0\\ 0 & \text{其他}\end{cases} \tag{9}$$

式中：M_{t+1}^{pr}为$t+1$时刻供电公司的售电利润；M_{t+1}^{co}为$t+1$时刻供电公司的补偿费用。

4)态势利导函数

在态势感知的基础上建立态势利导函数，自学习智能体通过环境理解函数的激励与惩罚实现决策优劣的训练学习，从而逐步实现自趋优决策。态势利导函数如式(10)所示。

$$L(\omega,b)=\sqrt{\frac{1}{m}\sum_{i=1}^{m}(Q_{\mathrm{tar}}^{(p)}(s_t,a_t)-Q_{\mathrm{pre}}^{(p)}(s_t,a_t,\omega,b))} \tag{10}$$

式中：p为控制策略；ω和b为DRL算法参数；m为经验池容量；$s_t\in S$为t时刻环境的状态；$a_t\in A$为t时刻CCR群组节点的响应状态。

$Q_{\mathrm{tar}}^{(p)}(s_t,a_t)$为样本训练后的预设值，其表达式如下：

$$Q_{\mathrm{tar}}^{(p)}(s_t,a_t)=Q_{\mathrm{pre}}^{(p)}(s_t,a_t,\omega,b)+\alpha\left(R_{t+1}+\gamma\max_{a_t\in A}Q_{\mathrm{pre}}^{(p)}(s_{t+1},a_t,\omega,b)-Q_{\mathrm{pre}}^{(p)}(s_t,a_t,\omega,b)\right) \tag{11}$$

式中：γ为折扣因子；α为学习率($0\leqslant\alpha\leqslant1$)。在基于DDQN的DRL算法中，$\alpha$的取值一般为[0.001,0.01]。

$Q_{\mathrm{pre}}^{(p)}(s_t,a_t,\omega,b)$为配电网数据导入后的计算值，其表达式为：

$$Q_{\mathrm{pre}}^{(p)}(s_t,a_t,\omega,b)=\mathcal{V}(s_t,\omega,b)+\left(\mathcal{A}(s_t,a_t,\omega,b)-\frac{1}{|A|}\sum_{a_t\in A}\mathcal{A}(s_t,a_t,\omega,b)\right) \tag{12}$$

$$\mathcal{V}(s_t,\omega,b)=\mathrm{Relu}(x\omega_0+b_0)\omega_1+b_1 \tag{13}$$

$$\mathcal{A}(s_t,a_t,\omega,b)=\mathrm{Relu}(x\omega_2+b_2)\omega_3+b_3 \tag{14}$$

式中：$|A|$为响应状态总数；$\mathrm{Relu}(x)=\max(0,x)$为线性整流函数；$\omega_0$为价值函数中与配电网状态相关的参数；$\omega_1$为价值函数中的结构参数；$\omega_2$为优势函数中与配电网状态相关的参数；$\omega_3$为优势函数中与响应状态相关的参数；$b_0$至$b_3$为偏置量。

3　多随机场景下CCR的优化调度决策

配电网随机运行方式下求解时，为适应多时间断面下的复杂场景，本章对历史负荷数据曲线进行泛化，基于泛化后的数据，通过时序差分法更新迭代预设值矩阵，利用ε-greedy策略选取最优动作，并引入经验池机制保证神经网络学习最新的观测状态。

1)负荷数据曲线泛化

本文在初始负荷的基础上，对非CCR群组节点，根据其节点峰谷功率差值进行叠加随机负荷，叠加基础值$\Delta P_{l,t}$如式(15)所示：

$$\Delta P_{l,t}=\frac{\Delta P_{l,\mathrm{d}}}{\sum\Delta P_{l,\mathrm{d}}}(P_{\mathrm{L},t}-P_{\mathrm{G},t})\qquad l\in N_{\mathrm{node}}\cap l\notin N_{\mathrm{CCR}} \tag{15}$$

式中：$\Delta P_{l,\mathrm{d}}$为节点l的峰谷功率差值；$P_{\mathrm{L},t}$为t时刻系统负荷需求；$P_{\mathrm{G},t}$为t时刻根节点输入功率；N_{node}为配电网的节点集合。

假设非CCR群组节点中节点π峰谷功率差值$\Delta P_{\pi,\mathrm{d}}$最大，将其作为平衡节点，其他非$\pi$节点且非CCR群组节点$l'$可叠加的功率$\Delta P'_{l',t}$如式(16)所示：

$$\Delta P'_{l',t}=\mu\frac{\Delta P_{l,t}}{\sum\Delta P_{l',\mathrm{d}}}\Delta P_{\pi,\mathrm{d}}\qquad l'\in N_{\mathrm{node}}\cap l'\notin N_{\mathrm{CCR}},l'\neq\pi \tag{16}$$

式中：$l'\in N_{\mathrm{node}}\cap l'\notin N_{\mathrm{CCR}},l'\neq\pi$；$\mu$为[−1,1]区间内的均匀分布值；$\Delta P_{l',\mathrm{d}}$为节点$l'$的峰谷功率差值。

负荷数据曲线泛化后，各非CCR群组节点功率如式(17)所示：

$$\begin{cases}P'_{l',t}=P_{l',t}+\Delta P'_{l',t}\\ P'_{\pi,t}=P_{\pi,t}-\sum\Delta P'_{l',t}\end{cases} \tag{17}$$

式中：$l'\in N_{\mathrm{node}}\cap l'\notin N_{\mathrm{CCR}},l'\neq\pi$；$P_{l',t}$和$P'_{l',t}$分别为泛化前、后节点$l'$在$t$时刻的功率；$P_{\pi,t}$和$P'_{\pi,t}$分别为泛化前、后节点$\pi$在$t$时刻的功率。

2)时序差分法机制

时序差分法搜索CCR群组优化调度策略如图1所示。阶段①初始状态s_1经过动作a_y至状态s_y，由式(5)计算R，并根据式(11)更新预设值矩阵，进入阶段②，并重复上述计算过程。基于Q-learning算法[23]，当已知优化响应状态空间与训练次数逐渐增大时，算法将逐步收敛，预设值矩阵迭代更新过程如式(18)所示。预设值及历史训练样本生成流程图如附录B图B1所示。

$$\boldsymbol{Q}=\begin{array}{c} \\ s_1 \\ \vdots \\ s_y \\ \vdots \\ s_z \end{array}\begin{array}{c} \begin{array}{ccccc} a_1 & \cdots & a_y & \cdots & a_z \end{array} \\ \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \end{array}\rightarrow\begin{array}{c} \begin{array}{ccccc} a_1 & \cdots & a_y & \cdots & a_z \end{array} \\ \begin{bmatrix} 0 & \cdots & Q_{1y} & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \end{array}\rightarrow\begin{array}{c} \begin{array}{ccccc} a_1 & \cdots & a_y & \cdots & a_z \end{array} \\ \begin{bmatrix} 0 & \cdots & Q_{1y} & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & Q_{yz} \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \end{array}\rightarrow\cdots \tag{18}$$

式中：Q_{1y}为式(18)迭代更新过程中在状态s_1下动作a_y对应的$Q_{\mathrm{pre}}^{(\rho)}(s_t,a_t,\omega,b)$的函数值，$Q_{yz}$的含义以此类推。

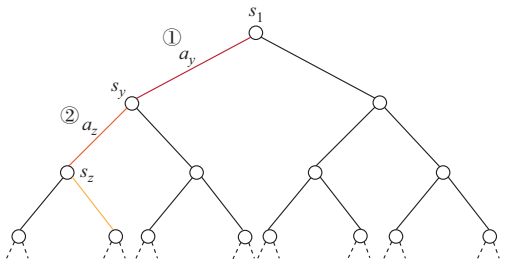

图1　时序差分法搜索机制

Fig. 1　Searching mechanism of temporal difference method

3)ε-greedy策略

训练过程中，学习初期随机选择动作从而积累观察样本，ε-greedy策略如式(19)所示：

$$a_t=\begin{cases}\text{random } A & \beta<\dfrac{T_{\mathrm{tr}}-t_{\mathrm{tr}}}{T_{\mathrm{tr}}}\varepsilon \\ \arg\max\limits_{a_t\in A} Q_{\mathrm{pre}}^{(\rho)}(s_t,a_t,\omega,b) & \beta\geqslant\dfrac{T_{\mathrm{tr}}-t_{\mathrm{tr}}}{T_{\mathrm{tr}}}\varepsilon\end{cases} \tag{19}$$

式中：random A表示从响应状态动作集A中随机选取动作；T_{tr}为训练总次数；t_{tr}为当前训练次数；β为[0,1]之间的随机数；ε为固定常数。

4)经验池设定

为了加快DRL算法训练速度与精确度，对经验池采取以下设定：

(1)经验池设置容量上限，从而消除样本采集时间接近而造成的强相关性。当产生样本数量超过经验池容量时，则剔除掉最早的观察样本再存入新样本。

(2)经验池设置观察值，当训练次数小于观察值时，不抽取训练样本。当经验池中样本数超过观察值时，则从中随机抽取小批量的观测样本，开展人工训练。

5)CCR群组优化调度策略求解

当观测状态由s_t变为s_{t+1}，进行以下3个判断步骤得到供电公司售电利润函数、响应状态函数以及电压运行回报函数的数值。首先，判断$U_{\mathrm{END},t+1}^{(i)}$是否大于$0.93U_{\mathrm{e}}$，根据式(6)计算电压运行回报函数$r_{i,t+1}^{\mathrm{u}}$的数值；然后，判断$M_{t+1}^{\mathrm{pr}}$是否小于$M_{t+1}^{\mathrm{co}}$，根据式(9)计算供电公司售电利润函数$r_{t+1}^{\mathrm{DSO}}$的数值；最后判断$(a_{t+1}^{(1)},a_{t+1}^{(2)},\cdots,a_{t+1}^{(n)})$是否等于$(a_t^{(1)},a_t^{(2)},\cdots,a_t^{(n)})$，根据式(7)和式(8)计算响应状态函数$r_{j,t+1}^{\mathrm{a}}$的数值。

根据式(5)计算R_{t+1}，更新预设值$Q_{\mathrm{tar}}^{(\rho)}(s_t,a_t)$，根据参数$\omega_0$至$\omega_3$及$b_0$至$b_3$更新计算值$Q_{\mathrm{pre}}^{(\rho)}(s_t,a_t,\omega,b)$，并放入经验池中，将新经验池中的$m$组$Q_{\mathrm{tar}}^{(\rho)}(s_t,a_t)$和$Q_{\mathrm{pre}}^{(\rho)}(s_t,a_t,\omega,b)$代入式(10)，得到态势利导函数$L(\omega,b)$，随后基于式(20)优化参数$\omega_0$至$\omega_3$及$b_0$至$b_3$，进入下一次迭代，重复上述过程直至$L(\omega,b)$收敛。在此过程中，状态集$A$的选取始终遵循ε-greedy策略。

$$\begin{cases}\omega_x=\omega_x-\alpha\dfrac{\partial L(\omega,b)}{\partial\omega_x} \\ b_x=b_x-\alpha\dfrac{\partial L(\omega,b)}{\partial b_x}\end{cases} \tag{20}$$

式中：$x=0,1,2,3$。

在高维数据场景下态势利导函数趋于收敛时，算法给出的CCR群组节点状态响应空间可被视为该组CCR群组节点在配电网调度下的最优状态空间。优化求解流程图如附录B图B2所示。

4　算例分析

4.1　随机场景

本文基于IEEE 33节点系统分析随机场景，如图2所示。算例分析将分别针对15 min采样间隔和30 min采样间隔进行优化策略求解，通过不同采样间隔形成不同规模的样本数量，验证所提方法的有效性。在图1中，节点17、21、24、32处安装电压量测装置，节点13、14、16、29、30以及31作为CCR群组节点与供电公司签订合同构成CCR群组，根节点及CCR群组节点安装功率量测装置。在日调度周期中，针对4个节点的电压，需要在64个状态空间中选择一组优化状态。当量测装置数据采样间隔为15 min时，日调度周期中存在96个时间断面，需在日周期内的2^{576}个样本数据中，从2^{576}个状态空间中进行策略优选。当数据采样间隔为30 min时，日调度周期中存在48个时间断面，需在日周期内的2^{288}个样本数据中，从2^{288}个状态空间中进行策略优选。

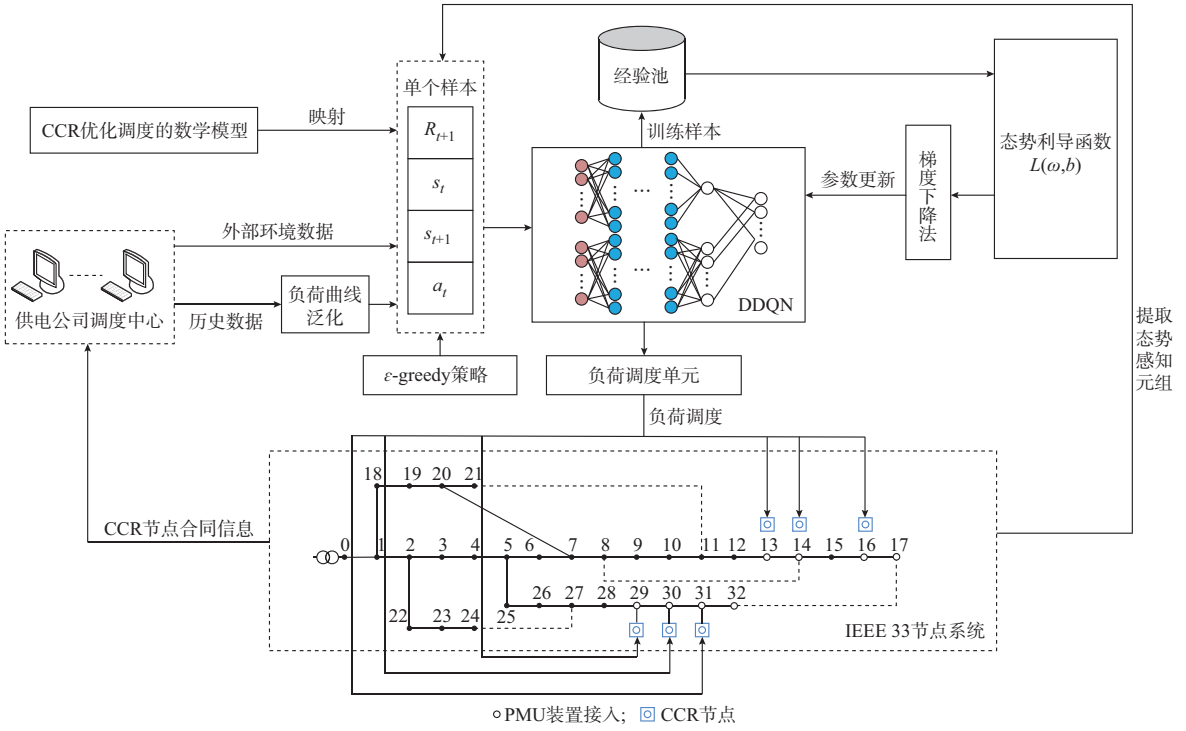

图2　基于DDQN结构的CCR群组节点的优化调度
Fig. 2　Optimal scheduling of nodes in CCR group based on DDQN structure

配电网的分时电价(购电和售电)以及所签订的合同内容分别见附录C表C1及表C2，CCR群组见表C3。为了尽量模拟用户用电的真实场景，体现用户负荷运行方式的多样性，算例模型中节点的实际日负荷曲线来源于IEEE欧洲低压试验馈线[24]。

4.2　算法参数及分析

1)算法参数

算法中态势感知元组参数设置如下：$F^{u}=0.5$；$F^{a}_{13}=F^{a}_{29}=-0.3$，$F^{a}_{14}=F^{a}_{30}=-0.2$，$F^{a}_{16}=F^{a}_{31}=-0.6$；$F^{p}=-0.4$。算法超参数选取如下：折扣因子$\gamma$取0.7，学习率$\alpha$取为0.007，$\varepsilon=1$，经验池容量为30，训练总次数$T_{tr}=3\,000$。

2)态势利导函数收敛分析

分别采用DDQN结构与DQN结构的DRL算法的态势利导函数衰减对比如附录D图D1所示。相比DQN结构，DDQN结构的态势利导函数衰减速度更快，衰减过程中波动更小，说明DDQN具有更优越的自学习能力。

3)学习率取值分析

学习率取值对比见附录D图D2。当学习率α为0.007时，态势利导函数收敛最快且收敛值最小，即此时DRL算法训练效果相对较优。

4.3　优选群组及优化策略分析

数据采样间隔为15 min的情况下，各CCR群组的计算值箱形图如图3所示，N5群组计算值最大，即为优选群组，该计算值对应的节点响应状态即为最优状态响应空间。

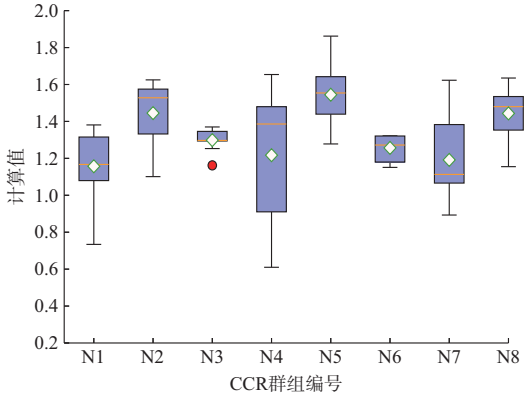

图3　N1至N8群组计算值箱形图
Fig. 3　Box-plot of calculated values for groups N1 to N8

针对N5群组基于DQN和DDQN的最优响应状态空间($a_t^{(13)}$，$a_t^{(16)}$，$a_t^{(29)}$，$a_t^{(31)}$)见表1。相对于基于DQN的最优决策，基于DDQN的最优决策累计停电次数更小。最优响应状态下N5群组节点核定削减负荷功率曲线如图4所示。

不同策略下的节点电压标幺值如表2所示，节点17、32的电压经过基于DDQN和DQN的DRL算法训练优化CCR群组节点的动作后，情况明显得到改善。

表1　基于DQN和DDQN的最优响应状态空间
Table 1　Optimal response state space based on DDQN and DQN

时间采样序列	响应状态空间		累计停电次数	
	DDQN	DQN	DDQN	DQN
71	(1,1,1,1)	(1,1,1,1)	0	0
72	(0,1,1,1)	(0,0,0,1)	1	3
74	(0,1,0,1)	(0,0,0,1)	2	3
75	(0,1,0,1)	(0,0,0,1)	2	3
76	(0,1,0,1)	(0,0,0,1)	2	3
77	(0,1,0,1)	(0,1,0,1)	2	3
78	(0,1,0,1)	(0,1,0,1)	2	3
79	(1,1,0,1)	(0,1,0,1)	2	3
80	(1,1,0,1)	(0,1,0,1)	2	3

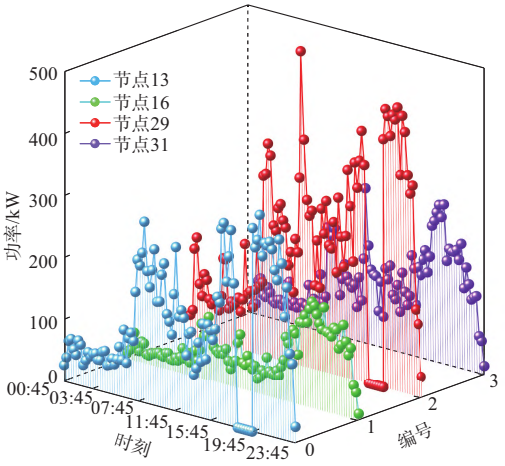

图4　最优响应状态下N5群组节点核定削减负荷功率曲线
Fig. 4　Approved load reduction power curve of group N5 nodes in optimal response state

表2　不同策略下的节点电压标幺值
Table 2　Per unit value of node voltage with different strategies

时间采样序列	节点17电压标幺值/p.u.			节点32电压标幺值/p.u.		
	初始值	DDQN	DQN	初始值	DDQN	DQN
72	0.923 0	0.932 5	0.938 0			
74	0.924 7	0.938 2	0.938 2			
75	0.923 9	0.941 3	0.941 3			
76	0.916 4	0.937 1	0.937 1	0.924 5	0.946 2	0.946 2
77	0.923 6	0.938 2	0.938 2	0.927 6	0.948 9	0.948 9
78	0.920 4	0.935 5	0.935 5	0.922 7	0.945 1	0.945 1
79	0.929 4	0.937 0	0.942 6			

供电公司在CCR群组节点的售电利润以及单组CCR的补偿见表3。相对基于DQN的最优决策结果，基于DDQN的最优决策CCR群组节点停电次数较少，改善了电压运行状态的同时，增大了供电公司的利润，减小了补偿费用。

表3　供电公司的售电利润以及CCR补偿费用
Table 3　Electricity sale profit of power supply company and CCR compensation cost

决策	售电利润/元	补偿费用/元
DDQN最优决策	26 421.250	5 555.125
DQN最优决策	19 173.325	12 803.050

采样间隔为30 min时，针对N5群组基于DDQN的最优响应状态空间($a_t^{(13)}$,$a_t^{(16)}$,$a_t^{(29)}$,$a_t^{(31)}$)见表4，节点电压标幺值如表5所示。由表4和表5可以看出，数据样本减少时策略仍然有效。

表4　基于DDQN的最优响应状态空间(30 min采样间隔)
Table 4　Optimal response state space based on DDQN (sampling interval of 30 min)

时间采样序列	响应状态空间	累计停电次数
35	(1,1,1,1)	0
36	(0,1,1,1)	1
37	(0,1,0,1)	2
38	(0,1,0,1)	2
39	(0,1,0,1)	2
40	(1,1,0,1)	2

表5　节点电压标幺值(30 min采样间隔)
Table 5　Per unit value of node voltage (sampling interval of 30 min)

时间采样序列	节点17电压标幺值/p.u.		节点32电压标幺值/p.u.	
	初始值	DDQN	初始值	DDQN
36	0.923 0	0.932 5		
37	0.924 7	0.938 2		
38	0.923 9	0.941 3		
39	0.916 4	0.937 1	0.924 5	0.946 2

5　结语

本文提出基于态势利导的需求响应自学习优化调度方法，实现了多随机场景下CCR群组的优选及对应节点的优化调度。主要工作如下：

1)针对需求响应的显著不确定性，本文基于MDP将其数学模型映射至态势感知元组，利用DRL算法自适应用户行为和电网运行状态的不确定性。

2)自学习智能体基于态势利导函数，通过环境理解函数的激励与惩罚实现决策优劣的训练学习，针对不同数量的数据样本实现了自趋优决策。

3)本文设置负荷数据曲线泛化机制、ε-greedy贪婪策略和经验池机制，针对多随机场景不同样本，分别在DQN和DDQN架构下开展自学习，验证了所提机制在随机复杂场景下的性能优越。

在双碳战略背景下,本文方法可为平抑规模化接入分布式能源带来的强随机性提供参考,下一步将深入开展用户侧可再生能源发电的随机性建模,探索新型电力系统需求侧响应随机优化运行的调度策略,为中国新型电力系统供需平衡、安全稳定运行提供技术保障。

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] 包铭磊,丁一,邵常政,等.北欧电力市场评述及对我国的经验借鉴[J].中国电机工程学报,2017,37(17):4881-4892.
BAO Minglei, DING Yi, SHAO Changzheng, et al. Review of Nordic electricity market and its suggestions for China [J]. Proceedings of the CSEE, 2017, 37(17): 4881-4892.

[2] 尹逊虎,丁一,惠红勋,等.初期现货市场下考虑用户响应行为的需求响应机制设计[J].电力系统自动化,2021,45(23):94-103.
YIN Xunhu, DING Yi, HUI Hongxun, et al. Design of demand response mechanism considering response behaviors of customers in initial electricity spot market[J]. Automation of Electric Power Systems, 2021, 45(23): 94-103.

[3] 郭昆健,高赐威,林国营,等.现货市场环境下售电商激励型需求响应优化策略[J].电力系统自动化,2020,44(15):28-35.
GUO Kunjian, GAO Ciwei, LIN Guoying, et al. Optimization strategy of incentive based demand response for electricity retailer in spot market environment[J]. Automation of Electric Power Systems, 2020, 44(15): 28-35.

[4] RAHIMI F, IPAKCHI A. Demand response as a market resource under the smart grid paradigm[J]. IEEE Transactions on Smart Grid, 2010, 1(1): 82-88.

[5] 赵鸿图,朱治中,于尔铿.电力市场中需求响应市场与需求响应项目研究[J].电网技术,2010,34(5):146-153.
ZHAO Hongtu, ZHU Zhizhong, YU Erkeng. Study on demand response markets and programs in electricity markets[J]. Power System Technology, 2010, 34(5): 146-153.

[6] 沈运帷,李扬,高赐威,等.需求响应在电力辅助服务市场中的应用[J].电力系统自动化,2017,41(22):151-161.
SHEN Yunwei, LI Yang, GAO Ciwei, et al. Application of demand response in ancillary service market[J]. Automation of Electric Power Systems, 2017, 41(22): 151-161.

[7] 郑若楠,李志浩,唐雅洁,等.考虑居民用户参与度不确定性的激励型需求响应模型与评估[J].电力系统自动化,2022,46(8):154-162.
ZHENG Ruonan, LI Zhihao, TANG Yajie, et al. Incentive demand response model and evaluation considering uncertainty of residential customer participation degree [J]. Automation of Electric Power Systems, 2022, 46(8): 154-162.

[8] 王韵楚,张智,卢峰,等.考虑用户行为不确定性的阶梯式需求响应激励机制[J/OL].电力系统自动化[2022-04-21].http://kns.cnki.net/kcms/detail/32.1180.TP.20220415.1651.008.html.
WANG Yunchu, ZHANG Zhi, LU Feng, et al. Stepwise incentive mechanism of demand response considering uncertainty of user behaviors[J/OL]. Automation of Electric Power Systems [2022-04-21]. http://kns. cnki. net/kcms/detail/32.1180. TP. 20220415.1651.008.html.

[9] 范明天,张祖平.电力系统优化数学模型和计算方法[M].北京:中国电力出版社,2013.
FAN Mingtian, ZHANG Zuping. Mathematical model and calculation method of power system optimization[M]. Beijing: China Electric Power Press, 2013.

[10] 张甜,赵奇,陈中,等.基于深度强化学习的家庭能量管理分层优化策略[J].电力系统自动化,2021,45(21):149-158.
ZHANG Tian, ZHAO Qi, CHEN Zhong, et al. Hierarchical optimization strategy for home energy management based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2021, 45(21): 149-158.

[11] 何清,李宁,罗文娟,等.大数据下的机器学习算法综述[J].模式识别与人工智能,2014,27(4):327-336.
HE Qing, LI Ning, LUO Wenjuan, et al. A survey of machine learning algorithms for big data[J]. Pattern Recognition and Artificial Intelligence, 2014, 27(4): 327-336.

[12] 刘全,翟建伟,章宗长,等.深度强化学习综述[J].计算机学报,2018,41(1):1-27.
LIU Quan, ZHAI Jianwei, ZHANG Zongzhang, et al. A survey on deep reinforcement learning[J]. Chinese Journal of Computers, 2018, 41(1): 1-27.

[13] 董瑶,葛莹莹,郭鸿湧,等.基于深度强化学习的移动机器人路径规划[J].计算机工程与应用,2019,55(13):15-19.
DONG Yao, GE Yingying, GUO Hongyong, et al. Path planning for mobile robot based on deep reinforcement learning [J]. Computer Engineering and Applications, 2019, 55(13): 15-19.

[14] 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.

[15] 夏伟,李慧云.基于深度强化学习的自动驾驶策略学习方法[J].集成技术,2017,6(3):29-34.
XIA Wei, LI Huiyun. Training method of automatic driving strategy based on deep reinforcement learning[J]. Journal of Integration Technology, 2017, 6(3): 29-34.

[16] 孙毅,刘迪,李彬,等.深度强化学习在需求响应中的应用[J].电力系统自动化,2019,43(5):183-191.
SUN Yi, LIU Di, LI Bin, et al. Application of deep reinforcement learning in demand response[J]. Automation of Electric Power Systems, 2019, 43(5): 183-191.

[17] XU H C, SUN H B, NIKOVSKI D, et al. Deep reinforcement learning for joint bidding and pricing of load serving entity[J]. IEEE Transactions on Smart Grid, 2019, 10(6): 6366-6375.

[18] XU H C, ZHANG K Q, ZHANG J B. Optimal joint bidding and pricing of profit-seeking load serving entity [J]. IEEE Transactions on Power Systems, 2018, 33(5): 5427-5436.

[19] 严干贵,阚天洋,杨玉龙,等.基于深度强化学习的分布式电采暖参与需求响应优化调度[J].电网技术,2020,44(11):4140-4149.
YAN Gangui, KAN Tianyang, YANG Yulong, et al. Demand response optimal scheduling for distributed electric heating based on deep reinforcement learning[J]. Power System Technology, 2020, 44(11): 4140-4149.

[20] HUANG X F, HONG S H, YU M M, et al. Demand response management for industrial facilities: a deep reinforcement learning approach[J]. IEEE Access, 7: 82194-82205.

[21] 徐弘升,陆继翔,杨志宏,等.基于深度强化学习的激励型需求响应决策优化模型[J].电力系统自动化,2021,45(14):97-103.
XU Hongsheng, LU Jixiang, YANG Zhihong, et al. Decision optimization model of incentive demand response based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2021, 45(14): 97-103.

[22] 李航,李国杰,汪可友.基于深度强化学习的电动汽车实时调度策略[J].电力系统自动化,2020,44(22):161-167.
LI Hang, LI Guojie, WANG Keyou. Real-time dispatch strategy for electric vehicles based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2020, 44(22): 161-167.

[23] WATKINS J, DAYAN P. Technical note: Q-learning [J]. Machine Learning, 1992, 8: 279-292.

[24] IEEE PES AMPS DSAS Test Feeder Working Group [EB/OL]. [2016-02-24]. http://sites. ieee. org/pes-testfeeders/resources/.

明威宇(1996—),男,硕士,主要研究方向:配电网规划运行、电力系统分析。E-mail:mwy1566@163.com

李 妍(1971—),女,通信作者,博士、副教授,硕士生导师,主要研究方向:电力系统运行与控制、配电网规划与评估、主动配电网技术等。E-mail:liyanhust@hust.edu.cn

程时杰(1945—),男,教授,博士生导师,中国科学院院士,IEEE Fellow,主要研究方向:人工智能在电力系统中的应用、电力系统运行与控制、超导电力等。E-mail:sjcheng@hust.edu.cn

(编辑 蔡静雯)

Self-learning Optimal Scheduling Method of Demand Response Based on Situation Orientation

MING Weiyu, LI Yan, CHENG Shijie, LONG Yu, XU Jing, WANG Shaorong
(State Key Laboratory of Advanced Electromagnetic Engineering and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: Aiming at the scene combination surge problem of the consumer choice resource (CCR) in multiple stochastic scenarios, this paper uses the deep reinforcement learning algorithm to achieve the optimal selection of CCR groups and the optimal scheduling of the contained nodes. First, according to the constraint conditions and objective function of optimal scheduling for CCR, the mathematical model and the solution complexity of the daily scheduling cycle are analyzed. Then, the optimal scheduling process for CCR is mapped into the situation awareness tuple based on the Markov decision process, and the situation orientation function is established based on the architecture of the dueling deep Q network. Through multiple situation deductions, the situation orientation function is derived by using the small batch gradient descent method, and the algorithm parameters are continuously fed back and updated to realize the decision optimization. Finally, based on the IEEE 33-bus example, by using random number of samples with different sizes, the optimization of the CCR group to be selected is realized in the random operation mode, and the corresponding optimal scheduling strategy is formulated.

This work is supported by Key Project of Smart Grid Technology and Equipment of National Key R&D Program of China (No. 2017YFB0902800).

Key words: consumer choice resource (CCR); deep reinforcement learning (DRL); dueling deep Q network (DDQN); Markov decision process (MDP); situation awareness; situation orientation