第46卷 第3期 2022年2月10日
DOI：10.7500/AEPS20201011005
电力系统自动化
Automation of Electric Power Systems
Vol. 46 No. 3 Feb. 10，2022

# 基于深度期望Q网络算法的微电网能量管理策略

冯昌森[1]，张 瑜[1]，文福拴[2]，叶承晋[2]，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 310007）

**摘要：**随着光伏发电在微电网中的渗透率不断提高，其发电出力的不确定性和时变性为微电网的经济运行带来了挑战。在构建经济调度模型时，就需要适当模拟不确定变量并相应地发展高效求解算法。在此背景下，文中提出能够有效计及不确定性因素的深度强化学习算法，以实时求解微电网的优化运行问题。首先，采用马尔可夫决策过程对微电网优化运行问题进行建模，用实时奖励函数代替目标函数和约束条件，利用其与环境互动，寻找最优策略。其次，借助贝叶斯神经网络对不确定的学习环境建模，进而在马尔可夫决策过程中有效考虑状态转移的随机过程。为此，提出双深度期望Q网络算法，通过考虑状态转移的随机性，优化一般深度Q网络算法的Q迭代规则，显著提高算法的收敛速度。最后，采用算例验证了所提模型和算法的有效性。

**关键词：**光伏发电；不确定性建模；深度强化学习；贝叶斯神经网络；双深度期望Q网络

## 0 引言

随着可再生能源发电技术的发展，光伏等分布式能源在电力系统中的渗透率不断提高，为电力系统安全和经济运行带来了挑战。受气候等环境因素的影响，光伏等分布式能源出力的不确定性为调度计划的制定带来了困难。如何对光伏出力的不确定性进行适当建模和高效求解是值得研究的重要问题。

在针对不确定性因素建模和求解方面已有相当多的研究成果。在不确定性建模方面，目前常用的方法主要有随机模型、模糊模型、区间数模型和机会约束模型。在随机模型方面，文献[1-2]用确定的分布描述光照强度的概率密度函数，其拟合效果受限于所选分布函数的种类。在区间数模型方面，文献[3]引入区间数来描述不确定性集合，规避极端条件下的风险，牺牲了系统运行的经济性。在机会约束模型方面，文献[4]将不确定性的调度模型转化成确定性优化问题，力图在最小化风险与最大化经济效益之间取得平衡。目前常用的求解不确定性优化模型方法包括混合整数规划[5]、动态规划[6]、随机线性规划[7]、改进微分进化算法[8]和飞蛾扑火算法[9]等。经典优化算法难以求得该类非线性优化模型的全局最优解，而启发式优化算法一般耗时很长。在此背景下，针对光伏发电高渗透率的微电网，需要对光伏发电出力进行更为精准的建模并寻求高效的求解算法。

深度强化学习通过与环境交互和反馈学习，不断改进策略并实现实时决策[10]，目前已经提出很多深度强化学习算法并将其应用于电力系统研究中。文献[11]利用深度Q网络（deep Q network，DQN）算法和文献[12]利用双深度Q网络（double deep Q network，DDQN）算法实现了配电网电压控制优化，DDQN算法避免了DQN算法中Q值被高估的问题。文献[13]利用对抗深度Q网络（dueling deep Q network，dueling-DQN）算法解决需求侧响应问题，dueling-DQN算法对DQN算法网络结构进行改进，提高了模型稳定性。文献[14]采用深度确定性策略梯度算法在连续动作空间中进行学习，实现微电网能源调度。上述强化学习算法根据决策主体与环境交互获得单一固定的状态转移过程，忽略了学习环境中不确定因素引起的状态转移的随机性[15]。贝叶斯神经网络提供概率建模方法，将贝叶斯神经网络与强化学习相结合，通过概率建模来估计学习环境的动态变化，并模拟马尔可夫决策过程中不确定环境下状态的随机转移[16]。与常规不确定性建模方法相比，深度学习方法可以更好地预测光伏发电出力的不确定性。同时，与长短期记忆网络[17]和深度置信神经网络[18]等确定性方法相比，贝叶斯神经网络[19]可以给出光伏发电功率的概率分布，提供更

收稿日期：2020-10-10；修回日期：2021-01-19。
上网日期：2021-04-23。
国家自然科学基金资助项目（51777193）；国家自然科学基金-山西煤基低碳联合基金项目（U1910216）。

为全面的预测信息，因而能更好地模拟光伏出力的不确定性。

在上述背景下，本文提出双深度期望Q网络（double deep expected Q network，DDEQN）算法，高效实时求解微电网随机经济调度问题。所提算法借助贝叶斯神经网络对强化学习中不确定的学习环境建模，进而优化DNQ算法中Q值的迭代规则，显著提高了神经网络的训练速度。

## 1　基于贝叶斯神经网络的光伏发电出力预测

### 1.1　贝叶斯神经网络

贝叶斯神经网络具有特殊的概率层结构，可以根据较小的数据量得到较为稳定的预测模型；同时，其概率层神经元的权重与偏置服从一定的概率分布，使其具备了描述不确定性变量的能力。

贝叶斯神经网络可以视为条件分布模型，用$X$、$Y$和$W$分别表示神经网络输入值、预测值和权重，则贝叶斯推理的原理可描述为：

$$p(W|X,Y)=\frac{p(W)p(Y|X,W)}{p(Y|X)} \tag{1}$$

式中：$p(W)$为$W$的先验概率；$p(Y|X,W)$为给定参数$W$和$X$下，输出$Y$的概率分布；$p(Y|X)$为给定参数$X$下输出$Y$的概率分布。

由于数据集已知，即$p(Y|X)$是常数，因此贝叶斯神经网络的原理可用式(2)描述，即利用给定参数下数据的概率分布$p(Y|X,W)$来计算后验概率$p(W|X,Y)$。

$$p(W|X,Y)\propto p(W)p(Y|X,W) \tag{2}$$

贝叶斯神经网络利用一个较为简单的分布，分布函数为$L(\mu,\sigma)$，采用变分推断的方法来逼近$p(W|X,Y)$，其中，$\mu$和$\sigma$为神经网络需要更新调整的参数。附录A描述了贝叶斯神经网络的具体训练算法。

### 1.2　光伏发电出力预测

基于贝叶斯神经网络的光伏出力预测，需要对多种影响因素进行分析。

1)决定性因素

光照辐射强度是影响光伏出力的决定性因素。光伏出力的表达式为：

$$P_{PV}=\phi D\eta \tag{3}$$

式中：$P_{PV}$为光伏发电出力；$\phi$为光照辐射强度；$D$为光伏阵列总面积；$\eta$为光电转换效率。

用独立的深度神经网络预测时段的天气类型、前一时段的光照辐射强度和某时间点的光伏出力最大值，以降低光伏预测模型的复杂度和训练难度，如附录B所示。

2)持续性影响因素

持续性影响因素指可在较长时间内对光伏出力产生影响的温度、相对湿度和风速等。可以从历史数据中挖掘其对光伏出力的影响。首先，计算温度、相对湿度和风速与光伏出力的皮尔逊系数以确定其相互依赖的定量关系。然后，通过对历史数据的学习，得到不同时移的持续性影响因素与光伏发电出力的相关系数。最后，通过深度全连接神经层将多维特征映射到低维中，在降低模型复杂度和提升训练效率的同时也保证了特征的完整性。

3)突发性影响因素

突发性影响因素指可在较短时间内对光伏出力产生影响，比如雾霾和运动云层等。将预测点的前一时刻的出力数据输入贝叶斯神经网络中，避免多时段输入造成的数据冗余。附录B详细介绍了各类因素建模过程、所采用的神经网络结构和基于贝叶斯神经网络预测光伏出力的流程图。

## 2　模型搭建

### 2.1　深度强化学习概述

强化学习问题可以建模为马尔可夫决策过程。马尔可夫决策过程主要包括5个元素：$<S,R,A,P,\gamma>$。其中，$S$为智能体感知到所有状态的集合；$A$为智能体可选择的所有动作的集合；$R$为智能体在状态$s$下采取动作$a$返回给智能体的奖励的集合；$P$为马尔可夫决策过程的状态转移概率，即在状态$s$下采取动作$a$转到下一状态$s'$的概率，可表示为$P_{a,ss'}$；$\gamma$为智能体下一状态与当前状态相关度的衰减因子。

强化学习用动作价值函数来评估动作选择的优劣，通过与环境的交互，不断更新迭代直至学习到最优策略。强化学习中Q学习算法得到广泛应用，学习主体基于状态$s$，采用$\varepsilon$-贪婪算法，即式(4)，选择动作$a$，得到奖励$r(s,a)$，进入状态$s'$后根据式(5)更新状态$s$下的价值函数$q_0(s,a)$。

$$r(s,a)=\begin{cases}1-\varepsilon & a=\arg\max\limits_{a\in A} q_0(s,a)\\ \varepsilon & a\neq\arg\max\limits_{a\in A} q_0(s,a)\end{cases} \tag{4}$$

$$q_0(s,a)=r(s,a)+\gamma\max_{a'\in A} q_0(s',a') \tag{5}$$

式中：$\varepsilon$为探索概率；$a'$为状态$s'$的动作。

为避免Q学习算法中的“维数灾”问题，深度强化学习被广泛研究与应用，即通过深度神经网络来拟合从输入状态到Q值的映射关系。例如，DQN算

法引入了经验回放机制，将与环境交互得到的奖励与状态更新情况保存起来，用于深度神经网络的训练。当神经网络参数收敛后，可以获得近似的Q值。

在DQN算法中，迭代后的Q值往往被高估，DDQN算法被进一步提出和研究。利用估计Q网络和目标Q网络来解耦动作的选择与目标Q值计算，估计Q网络与目标Q网络是2个结构相同但完全独立的网络，具体算法见文献[14]。

### 2.2　微电网能量管理模型

微电网能量管理中的可控设备包括储能系统、可控负荷和参与调度的电动汽车等。本文侧重研究基于深度强化学习的微电网随机调度的建模与求解，因而仅考虑储能系统作为可控资源。对于包含其他可控设备的场景，只需在本文模型的基础上改变马尔可夫决策过程中动作的维数即可。

以日运行成本最低为目标函数实现微电网的能量管理策略。日运行成本为调度周期内购电成本和储能系统运行成本之和，表达式为：

$$\min \sum_{t=1}^{T}\left(c_{\mathrm{b},t}|x_t|^{+}-c_{\mathrm{g},t}|-x_t|^{+}+\tau_t\right) \tag{6}$$

式中：$T$为调度时段数；$x_t$为时段$t$需要与主电网交换的电量，$x_t>0$表示从主电网购电，反之表示向主电网售电；$c_{\mathrm{b},t}$和$c_{\mathrm{g},t}$分别为在时段$t$从主电网购电的价格和向主电网售电的价格；$\tau_t$为时段$t$储能系统的运行成本；$|\cdot|^{+}$为取正运算。

同时，还需满足如下微电网运行约束。

1)功率平衡约束

$$x_t-P_{\mathrm{L},t}+P_{\mathrm{PV},t}-P_{\mathrm{ESS},t}=0 \tag{7}$$

式中：$P_{\mathrm{PV},t}$为时段$t$光伏的发电出力；$P_{\mathrm{ESS},t}$为时段$t$储能电池的功率，$P_{\mathrm{ESS},t}>0$表示储能系统充电，反之表示储能系统放电；$P_{\mathrm{L},t}$为时段$t$的负荷功率。

2)储能系统运行约束

$$\beta_{\min}<\beta_t<\beta_{\max} \tag{8}$$

$$0<P_{\mathrm{dis},t}<\bar{P}_{\mathrm{dis}} \tag{9}$$

$$0<P_{\mathrm{ch},t}<\bar{P}_{\mathrm{ch}} \tag{10}$$

$$P_{\mathrm{dis},t}P_{\mathrm{ch},t}=0 \tag{11}$$

$$\beta_{t+1}=\beta_t+\eta_{\mathrm{c}}P_{\mathrm{ch},t}\Delta t-\frac{P_{\mathrm{dis},t}}{\eta_{\mathrm{d}}}\Delta t \tag{12}$$

式中：$\beta_t$为储能系统在时段$t$的荷电状态；$\beta_{\min}$和$\beta_{\max}$分别为储能系统荷电状态允许的最小值和最大值；$P_{\mathrm{ch},t}$和$P_{\mathrm{dis},t}$分别为时段$t$储能系统的充、放电功率；$\eta_{\mathrm{c}}$和$\eta_{\mathrm{d}}$分别为储能系统的充、放电效率；$\bar{P}_{\mathrm{ch}}$和$\bar{P}_{\mathrm{dis}}$分别为储能系统充、放电功率的最大值；$\Delta t$为优化时段间隔。

受储能系统寿命和容量衰减的影响，在优化调度过程中需要考虑储能系统的度电成本[20-21]。度电成本是对储能系统全生命周期内的成本和发电量进行平准化后计算得到的储能成本。定义储能系统运行时的度电成本为$\lambda$，则时段$t$储能系统的运行成本表达式为：

$$\tau_t=\lambda|P_{\mathrm{ESS},t}| \tag{13}$$

3)调度周期内电池状态约束

$$\beta_0=\beta \tag{14}$$

式中：$\beta$为储能系统在调度周期末的荷电状态；$\beta_0$为储能系统在调度周期开始的荷电状态。

本文模型面向小型微电网，所有用电设备由同一条配电网馈线供电，地理位置较近，因而没有考虑电力潮流约束。且由于光伏出力为不确定性变量，上述模型目标函数应为期望值，相应的随机优化模型如式(15)所示，约束条件为式(8)—式(14)。

$$\min E\left(\sum_{t=1}^{T}\left(c_{\mathrm{b},t}|x_t|^{+}-c_{\mathrm{g},t}|-x_t|^{+}+\tau_t\right)\right) \tag{15}$$

式中：$E(\cdot)$为数学期望函数。

### 2.3　基于马尔可夫决策的能量管理模型

采用深度强化学习算法求解经济调度模型时，首先需要将式(15)建模为马尔可夫决策过程。

1)状态空间

状态即可观测变量。在构建状态空间时，既要考虑系统变量的多样性，又要考虑必要性，以保证系统的可观性。基于此，时段$t$的状态应该包括微电网中储能系统的荷电状态、实时负荷功率、实时光伏出力和下一时段光伏出力的预测值。光伏出力具有不确定性，本文模型中光伏下一时段出力由贝叶斯神经网络输出的概率分布表示。因此，时段$t$的状态空间$\Omega_{\mathrm{P}}$可表示为：

$$\Omega_{\mathrm{P}}=\{\beta_t,P_{\mathrm{PV},t},P_{\mathrm{L},t}\} \tag{16}$$

2)动作空间

动作即可调整变量。在本文模型中，通过储能系统的充放电和向电网买卖电量的动作来保证系统内部的功率平衡。其中，主网对微电网起支撑作用以保证微电网内部能量平衡，因此，无须在动作中表示，则时段$t$的动作空间可表示为储能设备功率挡位。

3)奖励

在深度强化学习中，将优化目标映射为奖励决策函数。时段$t$奖励$r_t$的表达式为：

$$r_t=c_{\mathrm{DW},t}+c_{\mathrm{ESS},t} \tag{17}$$

式中：$c_{\mathrm{DW},t}$为时段$t$的购电或售电奖励；$c_{\mathrm{ESS},t}$为时段$t$储能系统的运行奖励。

$$c_{\mathrm{DW},t}=\frac{c_{\mathrm{g},t}-c_{\mathrm{b},t}}{2}|x_t|-\frac{c_{\mathrm{g},t}+c_{\mathrm{b},t}}{2}x_t \tag{18}$$

储能系统的奖励包括其运行成本和违背运行约束的惩罚。违规惩罚是为了保证系统运行时满足相应的运行约束。本文模型的违规惩罚主要针对2个部分，即运行约束式(9)和式(14)。对于式(9)设置违规惩罚项$\upsilon_t$，表达式为：

$$\upsilon_t=-\delta|P_{\mathrm{ESS},t}| \tag{19}$$

式中：$\delta$为惩罚的单位成本。

在时段$t$，违规状态的表达式为：

$$\Delta\beta>\beta_{\max}-\beta_t \tag{20}$$

$$\Delta\beta>\beta_t-\beta_{\min} \tag{21}$$

$$\Delta\beta=\eta_{\mathrm{c}}|P_{\mathrm{ESS},t}|^{+}\Delta t+\left|-\frac{P_{\mathrm{ESS},t}}{\eta_{\mathrm{d}}}\right|^{+}\Delta t \tag{22}$$

式中：$\Delta\beta$为储能设备荷电状态的变化量。

对于约束(14)，周期末的储能系统状态与初始状态不相等时，设置一个较大的惩罚值$\Gamma$。

综上所述，时段$t$储能系统的运行奖励的表达式为：

$$c_{\mathrm{ESS},t}=\upsilon_t+\tau_t \tag{23}$$

这样，一个调度周期内的奖励$R$为：

$$R=\sum_{t=1}^{T}r_t+\Gamma \tag{24}$$

4)状态转移概率和折扣率

马尔可夫决策过程中，折扣率是对未来奖励的关注度，在算例计算时取固定值0.9。在状态$s$下选择动作$a$后，得到状态$s'$。根据式(16)可知，状态$s'$中的$\beta_{t+1}$可由式(12)获得，$P_{\mathrm{L},t+1}$为已知量，因此，状态转移概率可表示为$P_{\mathrm{PV},t+1}$的概率分布。

## 3　求解算法

### 3.1　DDEQN算法原理

无模型的强化学习算法根据智能体与环境交互获得单一固定的状态转移过程，忽略了学习环境中状态转移的随机问题。当强化学习的状态变量中包含不确定因素时，忽略状态的随机转移会影响深度强化学习算法的收敛速度，因此，本文提出了DDEQN算法。DDEQN算法将贝叶斯神经网络和深度强化学习结合起来，将状态转移的随机过程用贝叶斯神经网络表示，利用随机状态中动作价值函数期望值来更新Q网络。

首先，在估计Q网络中选择储能系统调度策略，表达式为：

$$a=\arg\max_{a\in A}q(s,a;\theta) \tag{25}$$

式中：$\theta$为估计Q网络的参数；$q(\cdot)$为Q网络的价值函数。

然后，在目标Q网络中更新价值函数，计算公式为：

$$q(s,a;\theta')=r(s,a)+\gamma E(q(s',a;\theta')) \tag{26}$$

式中：$\theta'$为目标Q网络的参数。

在时段$t$，贝叶斯神经网络预测下一时段的光伏出力，其概率密度函数为$\rho(s')$，则$E(q(s',a;\theta'))$可表示为：

$$E(q(s',a;\theta'))=\int_{-\infty}^{+\infty}q(s',a;\theta')\rho(s')\mathrm{d}s' \tag{27}$$

将概率密度函数离散化为：

$$E(q(s',a;\theta'))=\sum_{k=1}^{M}q(s'_k,a;\theta')p(s'_k) \tag{28}$$

式中：$M$为对贝叶斯神经网络的预测结果进行抽样并划分区间的总数；$s'_k$为第$k$个区间的状态。

这样，式(25)和式(26)可改写为：

$$a_k=\arg\max_{a_k\in A}q(s'_k,a_k;\theta) \tag{29}$$

$$q(s,a;\theta')=r(s,a)+\gamma\sum_{k=1}^{M}q(s'_k,a_k;\theta')p(s'_k) \tag{30}$$

式中：$a_k$为状态$s'_k$的动作。

DDEQN算法的流程图如图1所示。

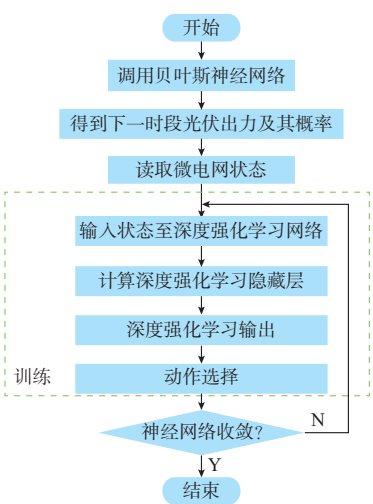


**图1　DDEQN算法流程图**

**Fig. 1　Flow chart of DDEQN algorithm**

### 3.2　DDEQN算法的神经网络训练参数设置

DDEQN算法的神经网络训练过程如附录C所示。神经网络训练过程中经验回放池、探索率和学习率都会影响神经网络的收敛性能，因此必须合理设定参数以保证神经网络学习过程收敛。

1)经验回放池

经验回放机制是为了避免经验数据的相关性，从以往的状态转移集合中随机采样进行训练。在训

练过程中,本文模型的动作集合较多,应该设置较大的经验回放池,以满足小批量训练时,随机抽样动作集合的多样性与全面性。

2)探索率

ε-贪婪算法中固定的ε会导致神经网络训练后期不收敛。本文设定ε随着训练次数逐渐减小来探索环境,以达到较好收敛效果。

3)学习率

学习率过大会导致过拟合现象,反之会使收敛速度变慢甚至停滞。因此必须通过多次尝试设定合适学习率。目标Q网络的参数由估计Q网络复制获得,因而也应设定合适的复制频率来避免过估计问题。

## 4 算例与结果

### 4.1 参数设置

以某小型产业园为微电网调度环境,某年5月到12月的光伏出力和园区总负荷为基础数据,对提出的模型与算法进行验证。该工业园区光伏出力、负荷功率和储能系统参数如附录D所示,实时电价引自文献[12],园区的储能系统度电成本引自文献[21]。

经过多次尝试,本文设定DDEQN算法中经验回放机制的样本存储量为4 800,每次小批量采样规模为600,初始探索率为0.1,最终探索率为0.001,探索步数为24 000,学习率取0.001,每训练10次更新一次目标Q网络参数。

本文使用Python语言并调用PyTorch包编写贝叶斯神经网络光伏发电出力预测程序,基于TensorFlow框架编写DDEQN算法程序。本文选用可以自适应改变学习率的自适应矩估计(adaptive moment estimation,ADAM)算法,具有更快的收敛速度和更好的收敛效果。计算机CPU型号为i7-8550U,内存为8 GB。贝叶斯神经网络训练步数为10 000,训练时长为22 h,DDEQN算法神经网络训练步数为70 000,训练时长为49 h。

### 4.2 贝叶斯神经网络训练结果

在贝叶斯神经网络中,用于特征提取的全连接层神经元数量为30,下一全连接层神经元数量为50,概率层神经元数量为55。选择7月10日(晴天)和9月6日(雨天)来验证预测结果。

由图2可知,晴天时,贝叶斯神经网络预测均值与实际值基本相等;雨天时,由于周围环境因素的复杂多变,贝叶斯神经网络的预测值在06:00时的误差较大,在其余时段仍然有较高精度。虽然预测精度与晴天相比有所下降,但预测值的变化符合实际出力的变化趋势,且预测误差在可接受范围内。

为进一步研究贝叶斯光伏预测的精度,将贝叶斯神经网络与全连接神经网络的预测结果进行对比,结果如附录B所示。贝叶斯神经网络具有更高精度和较小方差,这是因为基于贝叶斯神经网络的光伏出力预测有效考虑了不确定变量的时序相关性,从而提高了预测精度。本文中光伏出力的预测值取预测均值置信度为95%的区域,并以此在鲁棒性和经济性中取得平衡。

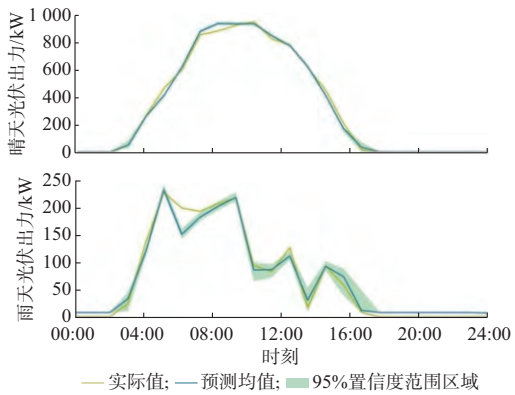


**图2 不同天气状况下光伏出力的预测结果与实际值对比**
**Fig. 2 Comparison of forecast results and actual values of photovoltaic generation output under different weather conditions**

### 4.3 仿真结果分析

为验证提出DDEQN算法的有效性,设计以下3种模式来对比分析。

模式1:采用DDQN算法,考虑光伏出力的不确定性,随机抽取贝叶斯神经网络的预测结果作为输入来训练深度神经网络。

模式2:采用DDEQN算法,即本文提出算法。

模式3:基于场景法的考虑光伏出力不确定性的随机优化算法。

选取一典型日园区光伏发电出力与负荷为基础数据(见附录D)。训练神经网络的过程中,采取两步训练法,在每一次训练后对神经网络进行测试,将一个调度周期中最优动作所对应的价值函数进行累加并归一化处理,用以表征神经网络的收敛程度,则第$i$步训练后神经网络的收敛率$\Theta(i)$为:

$$\Theta(i)=\frac{\sum_{t=1}^{T}q(s_{i,t},a;\theta_i)}{q^*} \tag{31}$$

式中:$s_{i,t}$为在时段$t$第$i$步的状态;$\theta_i$为第$i$步的参数;$q^*$为神经网络收敛后的累积$q$值。

为考察所提方法的收敛性能,还需预先确定$q^*$。经多次试验,模式1与模式2经过70 000步的训练

后均能收敛，故令 $q^*$ 为训练70 000步时的调度周期内最优动作所对应的累加值。

模式1与模式2的训练结果如图3所示，不同天气下的训练结果如附录E所示。

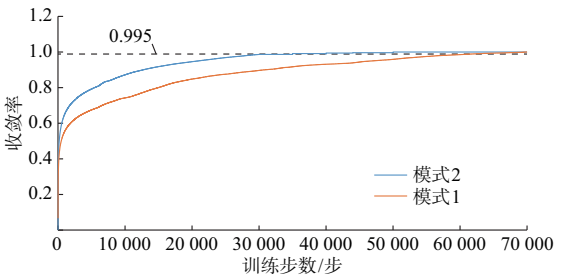


**图3　模式1与模式2的训练结果**
**Fig. 3　Training results of mode 1 and mode 2**

在收敛率达到0.995时，认为该神经网络收敛。由图3可知，模式2在训练约35 000步时收敛，而模式1在训练约67 000步时收敛。可见本文所提DDEQN算法具有更好的收敛性能。由于环境的复杂多变，为确保所提模型的精度与实用性，需要定期利用特殊的天气情况数据对模型进一步训练更新。

### 4.4　与随机优化算法的对比

采用4.2节的贝叶斯神经网络并用蒙特卡洛方法抽样10 000个场景作为随机优化模型中光伏出力的场景集合，优化模型如式(15)所示。设随机优化算法与深度强化学习算法中储能系统的充放电功率一致。3种模式下优化结果如图4所示。

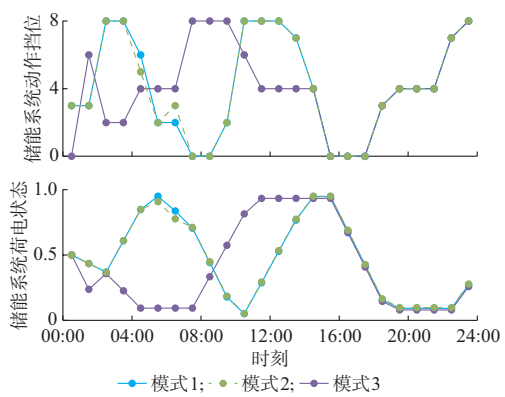


**图4　3种模式下的储能系统的荷电状态**
**Fig. 4　State-of-charge of battery storage system in three modes**

图4中，0至3代表储能系统放电的4个挡位，4代表储能系统不动作，5到8代表储能系统充电的4个挡位。由图4可知，模式3与模式1和2相比，在电价较低时段(01:00—06:00)和电价格较高的时段(08:00—10:00)的策略有较大差异。

为进一步对比深度强化学习算法与随机优化算法的优化结果，将得到的储能系统调度策略应用到测试场景集中，并将3种模式的优化结果进行对比，对比结果如表1所示。

**表1　不同模式的经济性对比**
**Table 1　Comparison of economy in different modes**

| 模式名称 | 成本/万元 | 光伏利用率/% |
|---|---|---|
| 模式1 | 189.07 | 81.33 |
| 模式2 | 184.30 | 81.57 |
| 模式3 | 215.04 | 76.26 |

由表1可知，相对于随机优化方法，深度强化学习方法的优化结果更具有经济性。此外，与传统DDQN算法相比，本文所提DDEQN算法的运行成本更低，这主要是因为DDEQN算法有更好的收敛性能。与随机优化模型相比，DDEQN算法基于贝叶斯神经网络对不确定性建模，考虑了不确定性变量在时序上的相关性，规避了极端小概率场景对整体策略的影响，因而具有较高的经济性。

### 4.5　模型继承能力评估

微电网的调度环境复杂多变，为了能快速获得不同环境下微电网的调度策略，可以借助迁移学习[22]的思想，对新环境下微电网的调度可以继承已有的神经网络并接续训练。本文定义在新环境下对现有神经网络模型继续训练的表现为继承能力。

为验证DDEQN算法训练得到的神经网络的继承能力，在新的微电网调度环境中设置光伏装机容量为原微电网的50%，储能系统的容量与各挡位充放电功率为原来的75%。设计模式4和模式5进行对比，模式4和模式5的继承能力评估如图5所示。

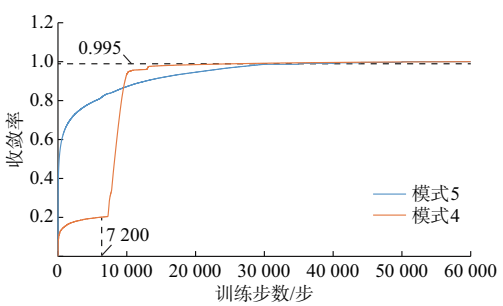


**图5　模式4和模式5的继承能力评估**
**Fig. 5　Evaluation of inheritance ability between mode 4 and mode 5**

模式4：用DDEQN算法重新进行训练。

模式5：利用模式2训练得到的神经网络继续训练。

由图5可知，与模式4相比，模式5具有更快的收敛速度。模式5继承已训练好的神经网络的结构

和参数,在训练步数大于观察步数(即7 200步)后,开始训练并迅速达到收敛,说明该神经网络具有较好的继承能力。

### 4.6 模型鲁棒性评估

为应对复杂多变的调度环境,有必要考察神经网络在不同调度环境下的性能表现,即鲁棒性。为评估本文模型的鲁棒性,分别针对贝叶斯神经网络和深度强化学习深度网络进行评估。

1)光伏出力预测的贝叶斯神经网络。用于光伏出力预测的数据为该区域1 a内周围环境数据。由4.2节可知,不同天气下基于贝叶斯神经网络的光伏出力预测结果均有较高的精度,说明贝叶斯神经网络具有较好的鲁棒性。在用新数据定期更新所提模型的基础上,日内调度环境发生变化时,所提模型仍具有良好的性能,即所提模型具有一定的实用性。

2)DDEQN算法的神经网络。为验证DDEQN算法的神经网络鲁棒性,基于4.5节新调度环境的参数设置,抽取60 d的光伏与负载数据为基础数据,设计模式6如下。

模式6:直接利用模式2训练得到的神经网络进行决策。

模式4与模式6的对比结果如附录E所示。通过对比可知本文所提的神经网络在新的调度环境中具有较好的鲁棒性。

### 4.7 不同结构神经网络的学习结果

神经网络的结构会影响训练过程与训练结果。本文对5种结构进行训练。5种结构的神经网络收敛性能如表2所示,收敛过程如附录E所示。

**表2 不同神经网络参数下的收敛性能**

**Table 2 Convergence performance with different neural network parameters**

| 结构名称 | 隐藏层数 | 单层神经元个数 | 神经元总数 | 收敛步数 | 收敛情况 |
|---|---|---|---|---|---|
| 结构1 | 2 | 55 | 110 | 56 000 | 慢 |
| 结构2 | 2 | 110 | 220 | 32 000 | 快 |
| 结构3 | 1 | 220 | 220 | 45 000 | 较慢 |
| 结构4 | 1 | 55 | 55 | | 欠拟合 |
| 结构5 | 3 | 550 | 1 650 | | 过拟合 |

由表2可知,结构2的收敛速度最快。结构1的隐藏层层数相同,但单层的神经元数量较少,导致其收敛速度较慢。结构3具有相同的神经元总数但隐藏层的层数不同,单层的神经元数量过多,降低了其收敛速度。结构4的神经元数量严重不足以致出现欠拟合现象,神经网络不能收敛。与结构4相反,结构5的神经元数量过多出现过拟合现象,在训练后期出现不稳定现象。

## 5 结语

本文提出了基于深度期望Q强化学习算法实现微电网实时经济调度。首先,采用贝叶斯神经网络对光伏出力不确定性建模,并在马尔可夫决策过程中适当考虑了状态转移的随机过程。然后,在传统无模型算法的基础上通过修改Q值的迭代规则提出了DDEQN算法,具有较好的收敛性能。最后,通过算例仿真对所提方法进行了验证,与现有方法相比,所提模型在经济性和收敛速度方面均有明显提升,且具有良好的继承能力与鲁棒性。此外,还对比分析了不同神经网络结构对训练过程和收敛性能的影响。

在后续的研究中,将进一步完善和发展所构建的模型,以形成广泛适用的方法,并用实际系统进行验证。

**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] 王蕾苹,卫志农,孙国强,等.计及分布式电源和负荷不确定性的多目标配网重构[J].电力自动化设备,2016,36(6):116-121.
WANG Xinping, WEI Zhinong, SUN Guoqiang, et al. Multi-objective distribution network reconfiguration considering uncertainties of distributed generation and load [J]. Electric Power Automation Equipment, 2016, 36(6): 116-121.

[2] 吴界辰,艾欣,胡俊杰,等.计及不确定因素的需求侧灵活性资源优化调度[J].电力系统自动化,2019,43(14):73-80.
WU Jiechen, AI Xin, HU Junjie, et al. Optimal dispatch of flexible resource on demand side considering uncertainties [J]. Automation of Electric Power Systems, 2019, 43(14): 73-80.

[3] 许寅,李佳旭,王颖,等.考虑光伏出力不确定性的园区配电网日前运行计划[J].电力自动化设备,2020,40(5):85-91.
XU Yin, LI Jiaxu, WANG Ying, et al. Day-ahead operation plan for campus distribution network considering uncertainty of photovoltaic output [J]. Electric Power Automation Equipment, 2020, 40(5): 85-91.

[4] 王健,谢桦,孙健.基于机会约束规划的主动配电网能量优化调度研究[J].电力系统保护与控制,2014,42(13):45-52.
WANG Jian, XIE Hua, SUN Jian. Study on energy dispatch strategy of active distribution network using chance-constrained programming [J]. Power System Protection and Control, 2014, 42(13): 45-52.

[5] 沈欣炜,郭庆来,许银亮,等.考虑多能负荷不确定性的区域综合能源系统鲁棒规划[J].电力系统自动化,2019,43(7):34-41.
SHEN Xinwei, GUO Qinglai, XU Yinliang, et al. Robust planning method for regional integrated energy system

considering multi-energy load uncertainties [J]. Automation of Electric Power Systems, 2019, 43(7): 34-41.

[6] 肖浩，裴玮，孔力.基于模型预测控制的微电网多时间尺度协调优化调度[J].电力系统自动化，2016，40(18)：7-14.
XIAO Hao, PEI Wei, KONG Li. Multi-time scale coordinated optimal dispatch of microgrid based on model predictive control [J]. Automation of Electric Power Systems, 2016, 40(18): 7-14.

[7] 朱嘉远，刘洋，许立雄，等.考虑风电消纳的热电联供型微电网日前鲁棒经济调度[J].电力系统自动化，2019，43(4)：40-48.
ZHU Jiayuan, LIU Yang, XU Lixiong, et al. Robust day-ahead economic dispatch of microgrid with combined heat and power system considering wind power accommodation [J]. Automation of Electric Power Systems, 2019, 43(4): 40-48.

[8] 张忠，王建学，曹晓宇.基于负荷分类调度的孤岛型微电网能量管理方法[J].电力系统自动化，2015，39(15)：17-23.
ZHANG Zhong, WANG Jianxue, CAO Xiaoyu. An energy management method of island microgrid based on load classification and scheduling [J]. Automation of Electric Power Systems, 2015, 39(15): 17-23.

[9] 潘晓杰，张立伟，张文朝，等.基于飞蛾扑火优化算法的多运行方式电力系统稳定器参数协调优化方法[J].电网技术，2020，44(8)：3038-3046.
PAN Xiaojie, ZHANG Liwei, ZHANG Wenchao, et al. Multi-operation mode PSS parameter coordination optimization method based on moth-flame optimization algorithm [J]. Power System Technology, 2020, 44(8): 3038-3046.

[10] 程乐峰，余涛，张孝顺，等.机器学习在能源与电力系统领域的应用和展望[J].电力系统自动化，2019，43(1)：15-31.
CHENG Lefeng, YU Tao, ZHANG Xiaoshun, et al. Machine learning for energy and electric power systems: state of the art and prospects [J]. Automation of Electric Power Systems, 2019, 43(1): 15-31.

[11] HUANG Q H, HUANG R K, HAO W T, et al. Adaptive power system emergency control using deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2020, 11(2): 1171-1182.

[12] YANG Q L, WANG G, SADEGHI A, et al. Two-time scale voltage control in distribution grids using deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2020, 11(3): 2313-2323.

[13] 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.

[14] 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.

[15] 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.

[16] NGUYEN T T, NGUYEN N D, NAHAVANDI S. Deep reinforcement learning for multiagent systems: a review of challenges, solutions, and applications [J]. IEEE Transactions on Cybernetics, 2020, 50(9): 3826-3839.

[17] 杨龙，吴红斌，丁明.新能源电网中考虑特征选择的Bi-LSTM网络短期负荷预测[J].电力系统自动化，2021，45(3)：166-173.
YANG Long, WU Hongbin, DING Ming. Short-term load forecasting in renewable energy grid based on bi-directional long short-term memory network considering feature selection [J]. Automation of Electric Power Systems, 2021, 45(3): 166-173.

[18] 李正明，梁彩霞，王满商.基于PSO-DBN神经网络的光伏短期发电出力预测[J].电力系统保护与控制，2020，48(8)：149-154.
LI Zhengming, LIANG Caixia, WANG Manshang. Short-term power generation output prediction based on a PSO-DBN neural network [J]. Power System Protection and Control, 2020, 48(8): 149-154.

[19] 赵康宁，蒲天骄，王新迎，等.基于改进贝叶斯神经网络的光伏出力概率预测[J].电网技术，2019，43(12)：4377-4386.
ZHAO Kangning, PU Tianjiao, WANG Xinying, et al. Probabilistic forecasting for photovoltaic power based on improved Bayesian neural network [J]. Power System Technology, 2019, 43(12): 4377-4386.

[20] FENG C S, WEN F S, SHI Y, et al. Coalitional game-based transactive energy management in local energy communities[J]. IEEE Transactions on Power Systems, 2020, 35(3): 1729-1740.

[21] 何颖源，陈永翀，刘勇，等.储能的度电成本和里程成本分析[J].电工电能新技术，2019，38(9)：1-10.
HE Yingyuan, CHEN Yongchong, LIU Yong, et al. Analysis of cost per kilowatt-hour and cost per mileage for energy storage technologies [J]. Advanced Technology of Electrical Engineering and Energy, 2019, 38(9): 1-10.

[22] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

冯昌森(1990—)，男，博士，讲师，主要研究方向：电力系统优化与控制、电力市场、机器学习等。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

文福拴(1965—)，男，教授，博士生导师，主要研究方向：电力系统故障诊断与系统恢复、电力市场与电力经济、智能电网与电动汽车等。E-mail：fushuan.wen@gmail.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、电能质量监控等。E-mail：youbingzhang@zjut.edu.cn

（编辑　杨松迎）

# Energy Management Strategy for Microgrid Based on Deep Expected Q Network Algorithm

*FENG Changsen*[1], *ZHANG Yu*[1], *WEN Fushuan*[2], *YE Chengjin*[2], *ZHANG Youbing*[1]

(1. College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China;
2. College of Electrical Engineering, Zhejiang University, Hangzhou 310007, China)

**Abstract:** With the increasing penetration of photovoltaic units in microgrids, the uncertainty and time-variability of its power generation have brought challenges to the economic operation of microgrids. When constructing an economic dispatch model, it is necessary to properly simulate the uncertain variables and develop the efficient solving algorithms accordingly. In this context, this paper proposes a deep reinforcement learning algorithm that can effectively account for uncertain factors to solve the problem of optimal operation of microgrids in real time. Firstly, the Markov decision process (MDP) is used to model the optimization operation of microgrids, and the real-time reward function is used to replace the objective functions and constraint conditions, and the interaction with the environment is used to find the optimal strategy. Secondly, the uncertain learning environment is modeled with the help of Bayesian neural networks, and then the stochastic process of state transfer is effectively considered in the MDP. Therefore, a double depth expected Q network algorithm is proposed. By considering the randomness of the state transfer, the Q iteration rules of the general deep Q network algorithm are optimized to significantly improve the convergence speed of the algorithm. Finally, a case is used to verify the effectiveness of the proposed model and algorithm.

This work is supported by National Natural Science Foundation of China (No. 51777193) and National Natural Science Foundation of China-Shanxi Coal-based Low Carbon Joint Fund (No. U1910216).

**Key words:** photovoltaic generation; uncertainty modeling; deep reinforcement learning; Bayesian neural network; double deep expected Q network