第45卷 第5期 2021年3月10日
DOI：10.7500/AEPS20200405004
电力系统自动化
Automation of Electric Power Systems
Vol. 45 No. 5 Mar. 10，2021

基于深度强化学习的综合能源系统动态经济调度

杨 挺，赵黎媛，刘亚闯，冯少康，盆海波
（智能电网教育部重点实验室（天津大学），天津市 300072）

摘要：综合能源系统的优化调度对于实现系统的多能互补和经济运行具有重要意义。然而，系统中可再生能源的间歇性以及用户用能需求的不确定性造成了系统中供需双方的随机波动，传统的调度方法难以准确地适应实际环境的动态变化。针对这一问题，提出了一种考虑可再生能源和负荷时变特性的综合能源系统动态经济调度方法。首先对综合能源系统动态经济调度问题进行数学描述，然后将该调度决策问题表述为强化学习框架，定义了系统的观测状态、调度动作和奖励函数，继而采用深度确定性策略梯度算法进行连续状态和动作空间下的动态调度决策。所提方法不需要对不确定性进行预测或建模，能够动态地对源和荷的随机波动做出响应。最后通过算例仿真验证了所提方法的有效性。

关键词：综合能源系统；动态经济调度；强化学习；深度确定性策略梯度

0 引言

随着环境压力的增加和可再生能源技术的发展，世界各国正调整能源结构，以减少对传统化石能源的依赖[1-2]。综合能源系统（integrated energy system）的构建为优化能源供应、提高能源效率提供了新的解决方案[3-4]。

综合能源系统的优化调度问题是综合能源系统研究的重要问题。针对系统经济调度问题，文献[5]研究了含热电联供（combined heat and power，CHP）机组、光伏（photovoltaic，PV）、风电机组等的微网经济调度问题，并采用Cplex软件进行求解。文献[6-7]采用改进的粒子群算法对电-热综合能源系统的经济调度模型进行求解。以上研究均基于可再生能源出力和用户负荷的准确预测信息，并未考虑源和荷的不确定性。为了应对系统中的不确定性，文献[8]采用场景分析法对风电、光伏出力随机性进行建模。文献[9]考虑热电联供型微网中负荷的不确定性，研究了基于鲁棒优化的系统优化调度问题。文献[10]采用不确定集合表征方法，以区间形式描述风速数据，构建了双层鲁棒模型，从而得到最恶劣场景下的系统调度方案。

收稿日期：2020-04-05；修回日期：2020-09-27。
上网日期：2020-12-16。
国家自然科学基金资助项目（61971305）；国家重点研发计划资助项目（2017YFE0132100）；天津市自然科学基金资助项目（19JCQNJC06000）。

上述文献主要研究综合能源系统的日前调度问题，多限于固定的调度计划，不能动态地对源和荷的随机变化做出响应。为解决上述问题，近年来模型预测控制[11]备受关注。文献[12]提出了一种基于模型预测控制的冷热电联供型微网的动态调度方法，以设备的日前计划出力为参考值，在日内调度中建立风电、光伏及负荷的预测模型，基于滚动优化求解出各设备的出力。文献[13]对并网型建筑能源系统采用模型预测控制方法优化各单元出力。虽然上述研究对综合能源系统的动态调度问题有很大贡献，但它们仍依赖于对可再生能源和负荷的精确预测。

本质上，综合能源系统的动态调度问题是随机序贯决策问题，可以采用强化学习（reinforcement learning，RL）进行求解。强化学习是一种重要的机器学习方法，它关注智能体在环境中如何采取行动以获得最大的累积回报[14]，而这与综合能源系统动态经济调度的设计目标是一致的，即关注综合能源系统如何进行调度决策以获得系统某个调度阶段最优的运行成本。为此，本文引入强化学习解决综合能源系统的动态经济调度问题。强化学习是一种无模型的方法，不依赖于不确定性的分布知识[15]，因此它不需要像传统方法那样预先对源和荷进行预测或建模。

已经有研究将强化学习用于电力能源系统的经济调度和能量管理中。文献[16]提出了一种基于多主体博弈和Q学习的综合能源微网协调调度方案。

为配合Q学习方法，该文将光伏、负荷需求等状态量及燃气联供单元出力等动作量进行了离散操作，但这样带来一个显著的问题就是维数灾难[17]。文献[18]研究了微网的分布式能量管理问题，将分布式电源、储能等建模为自治智能体，采用Q学习制定系统的能源管理和负荷调度策略，该文同样将柴油机和电储能(battery energy storage，BES)的动作进行了离散化处理。文献[19]采用深度Q网络(deep Q network，DQN)求解微网的实时调度策略，所提方法需要将电储能的充放电动作进行离散。然而，动作空间的离散化操作将大大减小可选动作范围。

为解决该问题，本文将所研究的综合能源系统动态调度问题置于连续状态和动作空间中进行处理，采用具有连续决策能力的深度确定性策略梯度(deep deterministic policy gradient，DDPG)算法[20]进行求解。

1 综合能源系统动态经济调度问题数学描述

综合能源系统运行优化的首要目标是提升系统经济效益，即在满足用户负荷需求的前提下，以最优经济运行为目标，有效地安排各设备在每个时段的出力。为此，本章建立了综合能源系统最优动态经济调度模型。以图1所示的综合能源系统为例，该系统包含了热电联供机组、光伏、电储能、燃气锅炉(gas boiler，GB)、电锅炉(electric boiler，EB)及用户电-热负荷等综合能源系统常见单元。

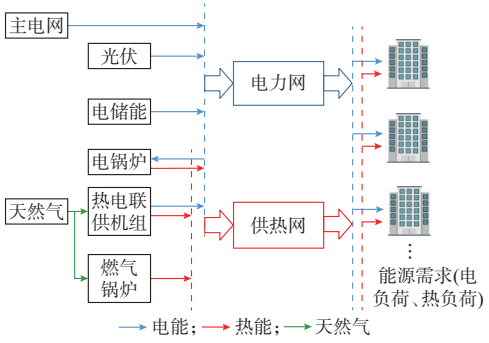

图1 综合能源系统结构示意图

Fig. 1 Schematic diagram of structure of integrated energy system

1.1 目标函数

综合能源系统动态经济调度问题的目标是最小化系统运行成本，其包括从能源供应处购买能源的成本、电储能的充放电折旧成本和设备维护成本。由于设备维护成本相对总运行成本较小，故未考虑在总成本中[21]。系统运行成本数学表示为：

$$F=\min\left(C_{\mathrm{E}}+C_{\mathrm{BES}}\right) \tag{1}$$

式中：C_{E}为购买能源的成本；C_{BES}为电储能的充放电折旧成本。

其中，购买能源的成本为：

$$C_{\mathrm{E}}=\sum_{t=1}^{T}\varepsilon_{\mathrm{e}}(t)\,p_{\mathrm{grid}}(t)\,\Delta t+\sum_{t=1}^{T}\varepsilon_{\mathrm{gas}}(t)\left(\frac{p_{\mathrm{CHP}}(t)\Delta t}{\eta_{\mathrm{CHPp}}}+\frac{h_{\mathrm{GB}}(t)\Delta t}{\eta_{\mathrm{GB}}}\right) \tag{2}$$

式中：$p_{\mathrm{grid}}(t)$为时段t系统与主电网进行电力交换的功率，为正表示系统向主电网购电，为负表示系统进行余电上网；$\varepsilon_{\mathrm{e}}(t)$为时段$t$的电价；$\varepsilon_{\mathrm{gas}}(t)$为时段$t$购买天然气的单位热值价格；$p_{\mathrm{CHP}}(t)$为时段$t$热电联供机组输出的电功率；$h_{\mathrm{GB}}(t)$为燃气锅炉输出的热功率；$\eta_{\mathrm{CHPp}}$为热电联供机组的电效率；$\eta_{\mathrm{GB}}$为燃气锅炉的效率；$T$为系统调度的总时段；$\Delta t$为时隙长度。

电储能的充放电折旧成本参考文献[22]中的计算公式得到：

$$C_{\mathrm{BES}}=\sum_{t=1}^{T}\rho_{\mathrm{BES}}\left|p_{\mathrm{BES}}(t)\right| \tag{3}$$

式中：$p_{\mathrm{BES}}(t)$为电储能在时段t的充电/放电功率，为正表示电储能处于放电状态，为负表示处于充电状态；ρ_{BES}为电储能折旧成本系数。

1.2 约束条件

综合能源系统动态经济调度问题的约束包括功率平衡约束、外部能源供应约束和设备运行约束。

1)功率平衡约束

在时段t，电功率平衡约束和热功率平衡约束分别可表述为：

$$p_{\mathrm{grid}}(t)+p_{\mathrm{PV}}(t)+p_{\mathrm{BES}}(t)+p_{\mathrm{CHP}}(t)-p_{\mathrm{EB}}(t)=p_{\mathrm{load}}(t) \tag{4}$$

$$h_{\mathrm{CHP}}(t)+h_{\mathrm{GB}}(t)+h_{\mathrm{EB}}(t)=h_{\mathrm{load}}(t) \tag{5}$$

式中：$h_{\mathrm{CHP}}(t)$为热电联供机组在时段t输出的热功率；$p_{\mathrm{EB}}(t)$为电锅炉的输入电功率；$h_{\mathrm{EB}}(t)$为电锅炉输出的热功率；$p_{\mathrm{PV}}(t)$为光伏的输出功率；$p_{\mathrm{load}}(t)$为时段t的电负荷；$h_{\mathrm{load}}(t)$为时段t的热负荷。

对于热电联供机组，其输出电功率与热功率之间的耦合关系称为“电热特性”，依据其热电比是否变化，可分为定热电比和变热电比2种类型。对热电联供机组，一般设为定热电比[10]，用变量b表示。

$$b=\frac{h_{\mathrm{CHP}}(t)}{p_{\mathrm{CHP}}(t)} \tag{6}$$

2)与主电网的交互功率约束

考虑到电网侧的运行稳定性，主网对综合能源系统的功率交互有上、下限约束要求：

$$P_{\mathrm{grid}}^{\min}\leqslant p_{\mathrm{grid}}(t)\leqslant P_{\mathrm{grid}}^{\max} \tag{7}$$

式中：$P_{\text{grid}}^{\max}$和$P_{\text{grid}}^{\min}$分别为系统与主电网交互功率的上限和下限。

3）设备运行约束

综合能源系统中各设备均有设备运行上、下限范围，对于热电联供机组输出电功率、电储能设备充电/放电功率、燃气锅炉输出热功率和电锅炉输出热功率，分别有

$$P_{\text{CHP}}^{\min} \leqslant p_{\text{CHP}}(t) \leqslant P_{\text{CHP}}^{\max} \tag{8}$$

$$P_{\text{BES}}^{\min} \leqslant p_{\text{BES}}(t) \leqslant P_{\text{BES}}^{\max} \tag{9}$$

$$H_{\text{GB}}^{\min} \leqslant h_{\text{GB}}(t) \leqslant H_{\text{GB}}^{\max} \tag{10}$$

$$H_{\text{EB}}^{\min} \leqslant h_{\text{EB}}(t) \leqslant H_{\text{EB}}^{\max} \tag{11}$$

式中：$P_{\text{CHP}}^{\min}$和$P_{\text{CHP}}^{\max}$分别为热电联供机组输出电功率的下限和上限；$P_{\text{BES}}^{\min}$和$P_{\text{BES}}^{\max}$分别为电储能充电/放电功率的下限和上限；$H_{\text{GB}}^{\min}$和$H_{\text{GB}}^{\max}$分别为燃气锅炉输出热功率的下限和上限；$H_{\text{EB}}^{\min}$和$H_{\text{EB}}^{\max}$分别为电锅炉输出热功率的下限和上限。

对于电储能设备，还需要避免深度充放电对电储能的损害，因此电储能的荷电状态（state of charge，SOC）被限定在一定范围内。

$$C_{\text{SOC}}^{\min} \leqslant c_{\text{SOC}}(t) \leqslant C_{\text{SOC}}^{\max} \tag{12}$$

式中：$C_{\text{SOC}}^{\min}$和$C_{\text{SOC}}^{\max}$分别为电储能荷电状态下限和上限；$c_{\text{SOC}}(t)$为电储能在时段t的荷电状态。

$c_{\text{SOC}}(t)$可表示为：

$$c_{\text{SOC}}(t) = c_{\text{SOC}}(t-1) - \frac{\eta_{\text{BES}} p_{\text{BES}}(t) \Delta t}{Q_{\text{BES}}} \tag{13}$$

$$c_{\text{SOC}}(0) = C_{\text{SOC}}^{\text{ini}} \tag{14}$$

式中：Q_{BES}为电储能的容量；$C_{\text{SOC}}^{\text{ini}}$为电储能初始时的荷电状态；$\eta_{\text{BES}}$为电储能充/放电系数，如式（15）所示。

$$\eta_{\text{BES}} = \begin{cases} \eta_{\text{ch}} & p_{\text{BES}} < 0 \\ \dfrac{1}{\eta_{\text{dis}}} & p_{\text{BES}} \geqslant 0 \end{cases} \tag{15}$$

式中：η_{ch}和η_{dis}分别为电储能的充电效率和放电效率。

此外，为保证电储能持续稳定运行，要求一个调度周期始末电储能容量相等。至此，综合考虑综合能源系统运行优化的目标如式（16）所示，系统所需满足的约束为式（4）—式（15）。

$$\min\left[\sum_{t=1}^{T}\varepsilon_{\text{e}}(t)\, p_{\text{grid}}(t)\Delta t + \sum_{t=1}^{T}\varepsilon_{\text{gas}}(t)\left(\frac{p_{\text{CHP}}(t)\Delta t}{\eta_{\text{CHPp}}} + \frac{h_{\text{GB}}(t)\Delta t}{\eta_{\text{GB}}}\right) + \sum_{t=1}^{T}\rho_{\text{BES}}\left|p_{\text{BES}}(t)\right|\right] \tag{16}$$

2　综合能源系统动态经济调度问题的强化学习框架

本文利用强化学习非常适合求解含不确定性因素的优化决策问题的优势，对计及间歇性可再生能源发电和用户负荷需求随机波动的综合能源系统的动态经济调度问题进行求解。首先，将第1章综合能源系统的动态经济调度问题的数学表述转化为强化学习框架。

强化学习的基本组成部分包括表征环境的状态集合S、表征智能体动作的动作集合A及对智能体的奖励r。在本文中，综合能源系统是智能体的环境，智能体通过调节系统中的设备出力进行最优调度决策。在时段t，环境向智能体提供观测到的系统状态$s_t \in S$，智能体基于策略π（策略π是将状态s映射到动作a的函数，即$\pi: S \to A \Rightarrow a = \pi(s)$）和综合能源系统状态$s_t$生成动态动作$a_t$。

这其中，综合能源系统的观测状态包括用户电负荷需求量、热负荷需求量、光伏发电功率、电储能的荷电状态以及所处的调度时段。对于综合能源系统，其状态表示为：

$$s_t = \{p_{\text{load}}(t), h_{\text{load}}(t), p_{\text{PV}}(t), c_{\text{SOC}}(t-1), t\} \tag{17}$$

在时段t，综合能源系统中的动作可以由设备的出力情况表示。由于$p_{\text{CHP}}(t)$确定后，$h_{\text{CHP}}(t)$可根据式（6）得到；$h_{\text{GB}}(t)$确定后，$h_{\text{EB}}(t)$可根据式（5）得到；进一步，$p_{\text{EB}}(t)$可以根据$p_{\text{EB}}(t) = h_{\text{EB}}(t)/\eta_{\text{EB}}$计算得到，其中$\eta_{\text{EB}}$是电锅炉的输出效率；进而$p_{\text{grid}}(t)$也可计算得到，故综合能源系统的动作可以用$p_{\text{CHP}}(t)$、$p_{\text{BES}}(t)$、$h_{\text{GB}}(t)$表示：

$$a_t = \{p_{\text{CHP}}(t), p_{\text{BES}}(t), h_{\text{GB}}(t)\} \tag{18}$$

综合能源系统经济调度的目标是最小化系统总运行成本。本文将系统总成本最小化问题转化为强化学习经典的奖励最大化形式，因此，将智能体在时段t获得的奖励表示为：

$$r_t(s_t, a_t) = -\frac{1}{1\,000}\left(C_{\text{E}}(s_t, a_t) + C_{\text{BES}}(s_t, a_t)\right) \tag{19}$$

式（19）中1/1 000是对成本值进行相应缩放。

在综合能源系统某一状态s_t确定时，综合能源系统动态经济调度动作a_t的优劣程度可以使用动作-值函数$Q_\pi(s,a)$[23]来评估，即

$$Q_\pi(s,a) = E_\pi\left(\sum_{k=0}^{T}\gamma^k r_{t+k}(s_{t+k}, a_{t+k})\Big|_{s_t=s_t, a_t=a_t}\right) \tag{20}$$

式中：$E_\pi(\cdot)$为策略π下的期望；$\gamma \in [0,1]$，为折扣因子，表示未来某一时刻的奖励在累积奖励中所占的影响比重，γ越大，则越重视对未来的奖励。

综合能源系统动态经济调度的目标是找到最优策略π^*以最大化动作-值函数，如式(21)所示。

$$\pi^* = \arg\max_{a\in A} Q_\pi(s,a) \tag{21}$$

3　综合能源系统动态经济调度问题求解方法

传统的强化学习方法在小规模离散空间的问题中表现良好。但当处理连续状态变量任务时随着空间维度的增加，其离散化得到的状态数量则呈指数级增长，即存在维数灾难问题[17]，无法有效学习。分析本文所研究的综合能源系统动态经济调度问题，其状态空间中的负荷、光伏发电及荷电状态均为连续量，因此传统强化学习方法往往无法有效求解。

同时，综合能源系统的动作空间中$p_{\mathrm{CHP}}(t)$、$p_{\mathrm{BES}}(t)$和$h_{\mathrm{EB}}(t)$也均为连续量。同样，对动作空间进行离散化将会删除决策动作域结构中的诸多信息。针对该问题，本文采用深度神经网络(deep neural network，DNN)[24]对强化学习进行函数近似，从而使其适用于连续状态和动作空间的综合能源系统动态经济调度问题。算法具体选择基于actor(策略)-critic(值)框架的DDPG算法[20]，它通过深度神经网络来估计最优策略函数。它不仅可以避免维数灾难，还能保存整个动作域的信息[25]。

DDPG算法使用2个独立的网络θ^Q和θ^π来逼近critic函数和actor函数，且每个网络均有各自的目标网络$\theta^{Q'}$和$\theta^{\pi'}$，其中Q'和π'分别为目标Q值和目标策略。

1)值网络训练

对于值网络，通过最小化损失函数$L(\theta^Q)$来优化参数：

$$L(\theta^Q) = E\left(y_t - Q\left(s_t, a_t \middle| \theta^Q\right)^2\right) \tag{22}$$

式中：y_t为目标Q值，如式(23)所示；$E(\cdot)$为期望函数。

$$y_t = r_t + \gamma Q'\left(s_{t+1}, \pi'\left(s_{t+1}\middle|\theta^{\pi'}\right)\middle|\theta^{Q'}\right) \tag{23}$$

在时段t，综合能源系统执行调度动作a_t后会进入下一个状态s_{t+1}，即更新后的电储能的荷电状态值和下一个时段观测得到的电负荷、热负荷和光伏发电值。

$L(\theta^Q)$关于θ^Q的梯度为：

$$\nabla_{\theta^Q} L(\theta^Q) = E\left(2\left(y_t - Q\left(s_t, a_t\middle|\theta^Q\right)\right)\nabla_{\theta^Q} Q(s_t, a_t)\right) \tag{24}$$

式中：∇为表示梯度计算的函数。

式(24)中$y_t - Q\left(s_t, a_t\middle|\theta^Q\right)$即为时序差分误差(timing differential error，TD-error)，根据梯度规则更新网络，可得到更新公式为：

$$\theta^Q \leftarrow \theta^Q - \mu_Q \nabla_{\theta^Q} L(\theta^Q) \tag{25}$$

式中：μ_Q为值网络学习率。

2)策略网络训练

对于策略网络，其提供梯度信息$\nabla_a Q\left(s_t, a_t\middle|\theta^Q\right)$作为动作改进的方向。为了更新策略网络，使用采样策略梯度：

$$\nabla_{\theta^\pi}\pi = \nabla_a Q\left(s, a\middle|\theta^Q\right)\Big|_{s=s_t, a=\pi(s_t)} \nabla_{\theta^\pi}\pi\left(s\middle|\theta^\pi\right)\Big|_{s=s_t} \tag{26}$$

根据确定性策略梯度，更新策略网络参数θ^π：

$$\theta^\pi \leftarrow \theta^\pi + \mu_\pi \nabla_{\theta^\pi}\pi \tag{27}$$

式中：μ_π为策略网络学习率。

目标网络的参数$\theta^{Q'}$和$\theta^{\pi'}$采用软更新技术来进一步提高学习过程的稳定性：

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'} \tag{28}$$

$$\theta^{\pi'} \leftarrow \tau\theta^\pi + (1-\tau)\theta^{\pi'} \tag{29}$$

式中：τ为软更新系数，$\tau \ll 1$。

算法中，通过为动作$a_t = \{p_{\mathrm{CHP}}(t), p_{\mathrm{BES}}(t), h_{\mathrm{GB}}(t)\}$加入随机噪声$\upsilon_t$以增加DDPG算法在综合能源系统交互时对环境的探索能力，以学习到更加优化的动态调度策略。

$$a_t = \pi\left(s_t\middle|\theta^\pi\right) + \upsilon_t \tag{30}$$

在此采用Ornstein-Uhlenbeck(OU)噪声。它是一种基于OU过程的随机变量，被用于模拟与时间关联噪声集[26]。

本文建立基于DDPG算法的综合能源系统动态经济调度框架，如图2所示。对于DDPG算法，策略网络的输入是5维状态$s_t = \{p_{\mathrm{load}}(t), h_{\mathrm{load}}(t), p_{\mathrm{PV}}(t), c_{\mathrm{SOC}}(t-1), t\}$，输出是3维动作$a_t = \{p_{\mathrm{CHP}}(t), p_{\mathrm{BES}}(t), h_{\mathrm{GB}}(t)\}$；值网络的输入是状态$s_t$和动作$a_t$，输出是动作-值函数，即$Q(s_t, a_t)$。在学习过程中，由于智能体与环境的顺序交互，样本是有关联的，这意味着这些样本并不像大多数深度学习算法所假设的那样是独立同分布的。为了应对此问题，DDPG算法采用了深度Q网络[27]中的经验回放机制。其通过在每个时段存储智能体的经验$e_t = (s_t, a_t, r_t, s_{t+1})$，形成回放记忆序列$D$。训练时，每次从$D$中随机提取小批量(mini-batch，大小为$M$)的经验样本，并基于梯度规则更新网络参数。经验回放机制通过随机采样历史数据打破了数据之间的相关性，而经验的重复使用也增加了数据的使用效率。

采用历史数据作为综合能源系统状态，离线训练DDPG算法网络。其输入为系统的电负荷、热负荷、光伏发电、电储能荷电状态及调度时段。离线训练结束后，训练得到的DDPG算法参数将被固定，用于综合能源系统的动态经济调度问题求解。对于综合能源系统，当调度任务来临时，在每个时段，根据当前系统状态s_t，利用训练好的DDPG算法网络、策略网络选择调度动作a_t。然后，执行动作a_t并且进入下一个环境状态，同时，获得奖励r_t。继而采集时段$t+1$系统的状态信息s_{t+1}作为新的样本，并进行这个时段的决策。如此，可以得到动态调度动作。

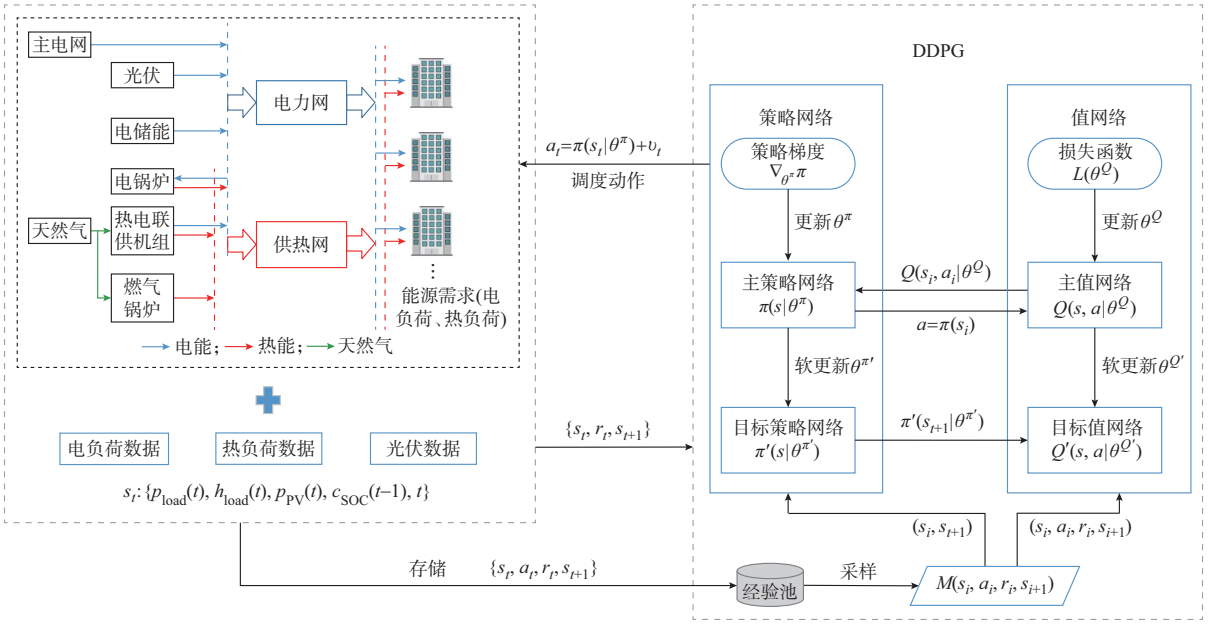

图2　基于DDPG算法的综合能源系统动态经济调度方案

Fig. 2　Dynamic economic dispatch scheme of integrated energy system based on DDPG algorithm

4　算例仿真与性能评价

为评测所提出的基于深度强化学习DDPG算法的综合能源系统动态调度策略的有效性，采用图1所示的热电联供型综合能源系统为算例进行仿真研究。系统中的热负荷、电负荷和同期光伏发电数据基于开源的CREST模型[28]产生。该模型是拉夫堡大学研究团队提出的，已经过有效性验证，且被广泛使用[29-31]。系统调度时段长度为24 h，相邻2个时段的间隔为15 min。综合能源系统中元件的运行参数如表1所示。

表1　设备运行参数

Table 1　Operation parameters of devices

设备类型	最小电(热)功率/MW	最大电(热)功率/MW
热电联供机组	0	1.6
电储能	−0.2	0.2
燃气锅炉	0	1.6
电锅炉	0	1.5

系统与主电网交换功率的范围为[−2.5, 2.5]MW，电储能的容量为1 000 kW·h，其他参数如表2所示。本文电价采用分时电价，如表3所示，其中峰时段为12：00—19：00，平时段为07：00—12：00、19：00—23：00，谷时段为23：00—07：00。天然气价格为固定价格0.4元/(kW·h)。

表2　其他参数

Table 2　Other parameters

参数	值	参数	值
η_{CHPp}	0.35	η_{dis}	0.95
b	1.20	ρ_{BES}	0.05
η_{GB}	0.80	C_{SOC}^{ini}	0.30
η_{EB}	0.95	C_{SOC}^{min}	0.20
η_{ch}	0.95	C_{SOC}^{max}	0.80

表3　分时电价

Table 3　Time-of-use electricity price

时段	购电电价/(元·(kW·h)$^{-1}$)	售电电价/(元·(kW·h)$^{-1}$)
峰时段	0.98	0.5
平时段	0.49	0.2
谷时段	0.17	0

4.1　训练过程

在将所建深度强化学习网络应用于系统动态经济调度问题之前，首先通过历史数据训练深度强化

学习的参数，得到深度强化学习网络。训练数据由去年同期、相同地点的负荷数据和光伏数据构造得到。在1日的开始，智能体接收来自环境的光伏出力、电负荷和热负荷需求，然后根据第3章所述的学习过程计算奖励值来调整DDPG算法网络参数，直到最终获得最大奖励。

以CREST模型生成的1月份和2月份的历史数据作为训练数据。图3给出了部分历史样本数据，包括电负荷、热负荷和光伏出力数据。对于超参数的选择，一方面根据深度学习社区[32]推荐的常用做法选取，另一方面，参考文献[23,33]中网络结构和参数的选择思路和方案，根据本文的训练数据进行试错调整。DDPG算法中策略网络和值网络的隐含层层数均为2层，每层有100个神经元，隐含层的激活函数均为ReLU(线性修正单元)。折扣因子为0.95，mini-batch大小为128，经验池大小为20 000，值网络学习率为0.001，策略网络学习率为0.000 1，τ为0.001，采用Adam优化器更新网络权重。

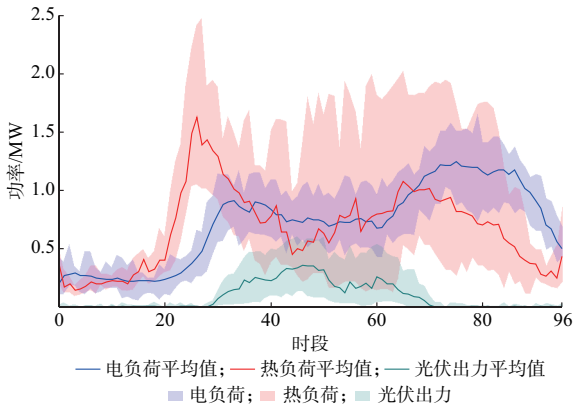

图3　综合能源系统的历史样本数据
Fig. 3　Historical sample data of integrated energy system

为了展现所提方法的收敛性能，附录A图A1给出了智能体训练过程中每100个episodes(周期)的平均奖励值曲线。该算法经过约10 000个episodes后收敛，得到了最优的动态经济调度策略。可以观察到，由于智能体最初对环境不熟悉，智能体执行调度决策后获得的奖励值较小。随着训练过程的继续，智能体不断地与环境交互并获得经验，因此奖励值整体趋势为逐渐增加并最终收敛。这说明智能体已经学习到了最小化系统运行成本的最优调度策略。由于在每个episode中的日训练数据，如负荷数据和光伏发电数据都有变化，因此在训练过程中奖励值会出现振荡。

4.2　动态调度结果

利用历史数据对DDPG算法网络进行离线训练后，得到的网络被保存以用于系统的动态经济调度。为了说明系统的动态经济调度结果，以该地区2016年2月1日的调度情况为例，基于本文提出的基于DDPG算法得到的$p_{CHP}(t)$、$p_{BES}(t)$、$h_{GB}(t)$以及由相应计算得到的$h_{CHP}(t)$、$h_{EB}(t)$、$p_{EB}(t)$、$p_{grid}(t)$结果如图4和图5所示。其中，图4为电功率调度结果，图5为热功率调度结果。

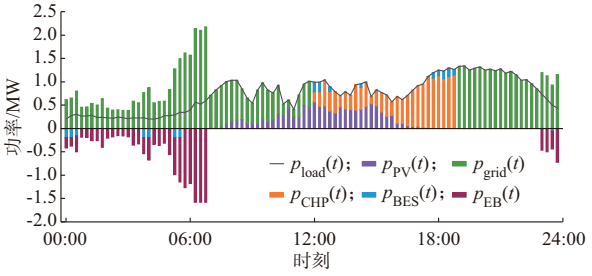

图4　基于DDPG算法的电功率调度结果
Fig. 4　Dispatch results of electric power based on DDPG algorithm

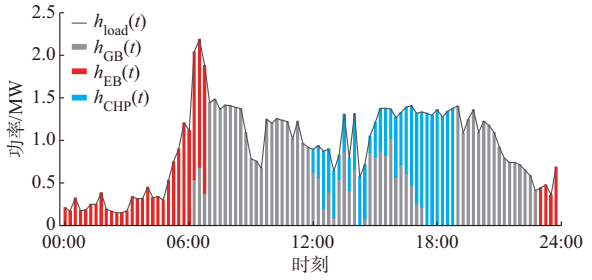

图5　基于DDPG算法的热功率调度结果
Fig. 5　Dispatch results of heat power based on DDPG algorithm

由图4可以看出，电储能在电价的引导下进行充放电，在谷电价且电负荷较小时充电以备后续的高峰时段，如00:00—00:30、03:45—04:00等时段；在峰电价且电负荷较高时放电以减少运行成本，如12:00—12:15、17:30—18:45等时段。在谷电价和平电价阶段，系统向主电网购电以满足用电需求。当电价为峰电价时，热电联供机组产生电能来避免向主电网购电，从而减少系统运行成本。由图5可以看出，电锅炉在谷电价时购电制热；谷电价且热负荷高于1.5 MW时，电锅炉制热不能满足热负荷需求，燃气锅炉制热进行补充，如06:15—06:45时段。平电价阶段，用户的热负荷需求由燃气锅炉制热满足。峰电价时热电联供机组或热电联供机组与燃气锅炉共同制热以满足用户热负荷需求。这表明本文提出的基于DDPG算法的动态经济调度方案能够不断优化各设备出力以满足用户负荷需求并减少系

统运行成本。

考虑到训练数据均为1月和2月（冬季）的样本数据，选取4月份（春季）的某天作为1个测试日，进一步验证所提方法的泛化性能。具体选择2016年4月9日的数据进行测试，该测试日的热负荷整体情况小于2016年2月1日测试日的热负荷，光伏出力时间较长。基于本文DDPG算法的调度方法得到的该测试日的调度结果如附录A图A2和图A3所示。

由图A2可以看出，电储能充放电情况和系统向主电网购电情况基本跟随电价变化。峰电价且光伏出力较大时，电负荷主要由光伏出力提供，不足部分由热电联供机组或电储能进行补充。由图A3可以看出，电锅炉在谷电价时购电制热；在07:00—07:30的平电价阶段，燃气锅炉制热不能满足热负荷需求，电锅炉产热进行补充。峰电价阶段，热负荷主要由热电联供机组或热电联供机组与燃气锅炉共同提供。在12:45、13:15—13:45等时刻或时段，热电联供机组未提供出力，热负荷全部由燃气锅炉提供。在2个测试场景中，所提方法均能够动态调整各设备出力以满足用户用能需求并降低系统运行成本，这说明所提方法对于未经历过的场景具有良好的泛化能力。

4.3　对比分析

为验证本文提出的基于DDPG算法的综合能源系统动态经济调度方法的有效性，将基于DDPG算法的调度方法与基于深度Q网络的调度方法以及基于模型预测控制的调度方法进行对比。从2016年1月和2月中随机抽取15 d作为15个测试日，对采用3种方法得到的运行成本进行比较。

对于所采用的深度Q网络，其输入为5维状态向量，输出为状态-动作对的Q值，本文将$p_{\mathrm{CHP}}(t)$、$p_{\mathrm{BES}}(t)$、$h_{\mathrm{GB}}(t)$分别以0.4、0.1、0.4 MW为间隔，分别离散为5个整数值。因此深度Q网络的输入层有5个神经元，输出层有125（=5×5×5）个神经元。深度Q网络有2个隐含层，每层均有200个神经元，隐含层的激活函数均为ReLU。对于模型预测控制方法，采用含1个隐含层的全连接神经网络对光伏出力和负荷进行预测。

表4给出了3种方法日运行成本的统计数据。其中，基于深度Q网络方法的平均日运行成本为17 928元，较DDPG算法增加了4.95%；基于模型预测控制方法的平均日运行成本为18 001元，较DDPG算法增加了5.37%。从日运行成本的平均值、最小值、最大值和标准差来看，DDPG算法较深度Q网络方法和模型预测控制方法获得了更好的性能，能有效地降低系统运行成本。由于源和荷的高度不确定性，采用基于源、荷预测信息的传统调度方法受限于预测准确度。而在深度Q网络方法中，由于燃气锅炉、热电联供机组、电储能的出力需取设定的离散值，而设定的离散动作值必将大大减少可行的动作选项，造成次优动作选择，因而造成了运行成本的增加。因此，本文所提出的基于DDPG算法的调度方法更适合解决综合能源系统的动态经济调度问题。

表4　不同调度方法的日运行成本统计
Table 4　Statistics of daily operational cost for different dispatch methods

方法	成本/元			
	平均值	最小值	最大值	标准差
DDPG算法	17 083	15 149	19 386	1 385
深度Q网络	17 928	15 543	20 837	1 597
模型预测控制	18 001	15 306	21 054	1 782

5　结语

本文提出了一种基于DDPG算法的综合能源系统动态经济调度方法。不同于传统方法，该方法不需要对源和荷进行预测，也不需要先期获得不确定性因素的分布知识。此外，所提方法通过将综合能源系统的动态调度问题置于连续状态和动作空间来处理，避免了离散化操作带来的维数灾难和次优调度策略选择问题。对比了2种不同的深度强化学习算法的应用效果，以及所提方法与传统方法的性能差异，并分析了差异产生的原因，表明了本文所提出的基于DDPG算法的综合能源系统动态调度方法能够更好地实现系统的动态经济调度。

在算法层面，深度强化学习DDPG算法的经验回放机制通过随机采样历史数据打破了数据之间的相关性，但该机制采用均匀随机采样的方式从经验池中提取经验数据，而没有考虑不同经验的重要程度。在今后的工作中，为了回放更有价值的经验，将对算法的经验回放机制进行改进，从而提高策略质量。本文所采用的深度强化学习方法亦可用于解决其他能源系统中的优化调度问题，如微网、能源互联网等，为这方面的研究提供了一种思路。

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] WANG C S, LV C, LI P, et al. Modeling and optimal

operation of community integrated energy systems: a case study from China[J]. Applied Energy, 2018, 230: 1242-1254.
[2] YU S W, HU X, LI L X, et al. Does the development of renewable energy promote carbon reduction? Evidence from Chinese provinces [J/OL]. Journal of Environmental Management, 2020, 268 [2020-03-20]. https://www.sciencedirect.com/science/article/pii/S0301479720305661.
[3] CHEN Z X, ZHANG Y J, TANG W H, et al. Generic modelling and optimal day-ahead dispatch of micro-energy system considering the price-based integrated demand response [J]. Energy, 2019, 176: 171-183.
[4] MASSRUR H R, NIKNAM T, AGHAEI J, et al. Fast decomposed energy flow in large-scale integrated electricity-gas-heat energy systems [J]. IEEE Transactions on Sustainable Energy, 2018, 9(4): 1565-1577.
[5] 李正茂,张峰,梁军,等.含电热联合系统的微电网运行优化[J].中国电机工程学报,2015,35(14):3569-3576.
LI Zhengmao, ZHANG Feng, LIANG Jun, et al. Optimization on microgrid with combined heat and power system [J]. Proceedings of the CSEE, 2015, 35(14): 3569-3576.
[6] 王丹,智云强,贾宏杰,等.基于多能源站协调的区域电力-热力系统日前经济调度[J].电力系统自动化,2018,42(13):59-67.
WANG Dan, ZHI Yunqiang, JIA Hongjie, et al. Day-ahead economic dispatch strategy of regional electricity-heating integrated energy system based on multiple energy stations [J]. Automation of Electric Power Systems, 2018, 42(13): 59-67.
[7] 刘洪,陈星屹,李吉峰,等.基于改进CPSO算法的区域电热综合能源系统经济调度[J].电力自动化设备,2017,37(6):193-200.
LIU Hong, CHEN Xingyi, LI Jifeng, et al. Economic dispatch based on improved CPSO algorithm for regional power-heat integrated energy system [J]. Electric Power Automation Equipment, 2017, 37(6): 193-200.
[8] 刘涤尘,马恒瑞,王波,等.含冷热电联供及储能的区域综合能源系统运行优化[J].电力系统自动化,2018,42(4):113-120.
LIU Dichen, MA Hengrui, WANG Bo, et al. Operation optimization of regional integrated energy system with CCHP and energy storage system [J]. Automation of Electric Power Systems, 2018, 42(4): 113-120.
[9] WANG R, WANG P, XIAO G X. A robust optimization approach for energy generation scheduling in microgrids [J]. Energy Conversion and Management, 2015, 106: 597-607.
[10] 朱嘉远,刘洋,许立雄,等.考虑风电消纳的热电联供型微网日前鲁棒经济调度[J].电力系统自动化,2019,43(4):40-51.
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): 40-51.
[11] JIN L C, KUMAR R, ELIA N. Model predictive control-based real-time power system protection schemes [J]. IEEE Transactions on Power Systems, 2010, 25(2): 988-998.
[12] 吴鸣,骆钊,季宇,等.基于模型预测控制的冷热电联供型微网动态优化调度[J].中国电机工程学报,2017,37(24):7174-7184.
WU Ming, LUO Zhao, JI Yu, et al. Optimal dynamic dispatch for combined cooling heating and power microgrid based on model predictive control[J]. Proceedings of the CSEE, 2017, 37(24): 7174-7184.
[13] ZHAO Y, LU Y H, YAN C C, et al. MPC-based optimal scheduling of grid-connected low energy buildings with thermal energy storages[J]. Energy and Buildings, 2015, 86: 415-426.
[14] SUTTON R S, BARTO A G. Introduction to reinforcement learning[M]. Cambridge, USA: MIT Press, 1998.
[15] RUELENS F, CLAESSENS B J, VANDAEL S, et al. Residential demand response of thermostatically controlled loads using batch reinforcement learning [J]. IEEE Transactions on Smart Grid, 2017, 8(5): 2149-2159.
[16] 刘洪,李吉峰,葛少云,等.基于多主体博弈与强化学习的并网型综合能源微网协调调度[J].电力系统自动化,2019,43(1):40-50.
LIU Hong, LI Jifeng, GE Shaoyun, et al. Coordinated scheduling of grid-connected integrated energy microgrid based on multi-agent game and reinforcement learning[J]. Automation of Electric Power Systems, 2019, 43(1): 40-50.
[17] YANG T, ZHAO L Y, LI W, et al. Reinforcement learning in sustainable energy and electric systems: a survey[J]. Annual Reviews in Control, 2020, 49: 145-163.
[18] FORUZAN E, SOH L K, ASGARPOOR S. Reinforcement learning approach for optimal distributed energy management in a microgrid[J]. IEEE Transactions on Power Systems, 2018, 33(5): 5749-5758.
[19] JI Y, WANG J H, XU J C, et al. Real-time energy management of a microgrid using deep reinforcement learning [J/OL]. Energies, 2019, 12 (12) [2020-03-20]. https://www.mdpi.com/1996-1073/12/12/22912291.
[20] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[EB/OL]. [2020-03-20]. https://arxiv.org /abs/1509.02971v2.
[21] MOSLEHI S, REDDY T A. A new quantitative life cycle sustainability assessment framework: application to integrated energy systems[J]. Applied Energy, 2019, 239: 482-493.
[22] YU L, XIE W W, XIE D, et al. Deep reinforcement learning for smart home energy management [J]. IEEE Internet of Things Journal, 2020, 7(4): 2751-2762.
[23] WAN Z Q, LI H P, HE H B, et al. Model-free real-time EV charging scheduling based on deep reinforcement learning [J]. IEEE Transactions on Smart Grid, 2019, 10(5): 5246-5257.
[24] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.
[25] YE Y J, QIU D W, SUN M Y, et al. Deep reinforcement learning for strategic bidding in electricity markets [J]. IEEE Transactions on Smart Grid, 2020, 11(2): 1343-1355.
[26] UHLENBECK G E, ORNSTEIN L S. On the theory of the Brownian motion[J]. Physical Review, 1930, 36(5): 823-832.
[27] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518(7540): 529-533.
[28] MCKENNA E, THOMSON M. High-resolution stochastic integrated thermal-electrical domestic demand model [J]. Applied Energy, 2016, 165: 445-461.
[29] GAN L K, HUSSAIN A, HOWEY D A, et al. Limitations

in energy management systems: a case study for resilient interconnected microgrids [J]. IEEE Transactions on Smart Grid, 2019, 10(5): 5675-5685.
[30] THOMAS D, D'HOOP G, DEBLECKER O, et al. An integrated tool for optimal energy scheduling and power quality improvement of a microgrid under multiple demand response schemes [J/OL]. Applied Energy, 2020, 260 [2020-03-20]. https://www.sciencedirect.com/science/article/abs/pii/S030626191932001X.
[31] AYYADI S, BILIL H, MAAROUFI M. Optimal charging of electric vehicles in residential area [J/OL]. Sustainable Energy, Grids and Networks, 2019, 19 [2020-03-20]. https://www.sciencedirect.com/science/article/abs/pii/S2352467719300761.
[32] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep learning [M]. Cambridge, USA: MIT Press, 2016.
[33] 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)：33-46.
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): 33-46.

杨　挺(1979—)，男，通信作者，博士，教授，主要研究方向：人工智能与大数据、电力信息物理系统。E-mail: yangting@tju.edu.cn

赵黎媛(1992—)，女，博士研究生，主要研究方向：人工智能、综合能源系统能量管理。E-mail: yuaner_zhao@tju.edu.cn

刘亚闯(1989—)，男，博士研究生，主要研究方向：电力信息物理系统优化控制。E-mail: liuyachuang@foxmail.com

（编辑　顾晓荣）

Dynamic Economic Dispatch for Integrated Energy System Based on Deep Reinforcement Learning

YANG Ting, *ZHAO Liyuan*, *LIU Yachuang*, *FENG Shaokang*, *PEN Haibo*

(Key Laboratory of the Ministry of Education on Smart Power Grids (Tianjin University), Tianjin 300072, China)

Abstract: The optimal dispatch of integrated energy systems is of great significance for the realization of multi-energy complementary and economic operation of the system. However, the intermittence of renewable energy and the uncertainty of users' energy demands in the system cause the random fluctuation on both the supply and demand sides in the system. Traditional dispatch methods are difficult to adapt to the dynamic changes of the actual environment accurately. In accordance to this problem, a dynamic economic dispatch method for integrated energy systems considering the time-varying characteristics of renewable energy and heterogeneous loads is proposed. Firstly, the dynamic economic dispatch problem for integrated energy systems is described mathematically. Secondly, the dispatch decision problem is formulated as a reinforcement learning framework, in which the observation state, dispatch action and reward function of the system are defined. Then, deep deterministic policy gradient (DDPG) algorithm is used to make dynamic dispatch decisions in continuous state and action spaces. The proposed method does not need to predict or model the uncertainty, and can dynamically respond to the random fluctuations of the source and loads. Finally, simulation is carried out to demonstrate the effectiveness of the proposed method.

This work is supported by National Natural Science Foundation of China (No. 61971305), National Key R&D Program of China (No. 2017YFE0132100), and Tianjin Municipal Natural Science Foundation of China (No. 19JCQNJC06000).

Key words: integrated energy system; dynamic economic dispatch; reinforcement learning; deep deterministic policy gradient