

基于二阶随机动力学的多虚拟电厂 自趋优能量管理策略

陈嘉琛, 陈中*, 李冰融, 刘汶瑜, 潘俊迪

(东南大学电气工程学院, 江苏省南京市 210018)

Energy Management Strategy for Multi-virtual Power Plants With Self-optimization Based on Second-order Stochastic Dynamics

CHEN Jiachen, CHEN Zhong*, LI Bingrong, LIU Wenyu, PAN Jundi¹

(School of Electrical Engineering, Southeast University, Nanjing 210018, Jiangsu Province, China)

ABSTRACT: The presence of numerous stochastic elements in distributed energy resources (DERs) leads to frequent changes in Multi-Virtual Power Plant (MVPP) when it comes to the strategy of individual VPPs. For a given entity, understanding the trend of the impact on its own returns when perceiving sudden changes in the strategies of other entities and rapidly adjusting its own optimization strategy is a critical issue that urgently needs to be addressed. This paper proposes a self-trending optimization strategy for MVPPs based on second-order stochastic dynamics, aiming to enhance the autonomy of VPPs in responding to changes in the strategies of other entities. First, addressing the heterogeneous operational characteristics of DERs, the paper focuses on the adjustable space of resources to construct a clustered operational model for VPP resources. Next, the stochastic nature of VPP strategy transitions is depicted based on the theory of random graphs. Then, second-order stochastic dynamic equations are used to explore its spontaneous evolutionary information to adjust the comprehensive profit of VPPs with the change of other entities' strategies. Moreover, the adjusted profit is used as the true reward function for the Integrated Soft Actor-Critic (ISAC) deep reinforcement learning decision model to establish a multi-agent distributed solution framework. Finally, multiple algorithm comparison experiments are designed to validate the self-trending performance of the proposed strategy in this paper.

KEY WORDS: multi-virtual power plant; self-optimization; aggregate operation model; second-order stochastic dynamics; multi-agent deep reinforcement learning

基金项目: 国家自然科学基金项目(52077035)。

Project Supported by National Natural Science Foundation of China (52077035).

摘要: 分布式资源(distributed energy resources, DERs)的随机元素会引起多虚拟电厂(multi-virtual power plant, MVPP)系统内虚拟电厂(virtual power plant, VPP)策略频繁变化。对于某主体,如何感知其他主体策略突然变化时对自身收益的影响趋势,并快速调整自身策略,是亟需解决的难点。该文提出基于二阶随机动力学的多虚拟电厂自趋优能量管理策略,旨在提升VPP应对其他主体策略变化时的自治能力。首先,针对DERs异质运行特性,聚焦可调空间构建VPP聚合运行模型;然后,基于随机图描绘VPP策略变化的随机特性;其次,用二阶随机动力学方程(stochastic dynamic equation, SDE)探索VPP收益结构的自发演化信息,修正其他主体策略变化时自身综合收益;再次,将修正收益作为融合软动作-评价(integrated soft actor-critic, ISAC)强化学习算法的奖励搭建多智能体求解框架。最后,设计多算法对比实验,验证了该文策略的自趋优性能。

关键词: 多虚拟电厂; 自趋优; 聚合运行模型; 二阶随机动力学; 多智能体强化学习

0 引言

虚拟电厂(virtual power plant, VPP)正趋向于成为由运营商调控的含规模化异质分布式资源(distributed energy resources, DERs)的集群实体^[1-2]。多虚拟电厂(multi-virtual power plant, MVPP)的高效协同能量管理可以有效提高可再生能源的消纳水平。然而,单个VPP策略突然变化给MVPP统一聚合调度带来了压力^[3]。如何挖掘VPP应对其他主体策略突然变化时的自趋优能力是MVPP研究中亟待解决的难题之一。

当前诸多研究对异质分布式资源采用传统的逐一建模方式。文献[4]考虑新能源发电站、传统热发电站与碳捕集设备组建成VPP联合运行;文献[5]

深入挖掘可控负荷、电动汽车等分布式资源 (distributed energy resources, DERs) 的调度潜能, 建立了虚拟电厂日前竞标模型。然而, 上述文献仅对特定种类的 DERs 进行建模, 无法应对异质运行特性。为此, 一些研究尝试用可调空间作为建模基准构建 DERs 聚合运行模型。文献[6]分别建立储能型和非储能型 DERs 的统一模型; 文献[7]建立半控型和可控型 DERs 的统一模型, 研究 VPP 参与调峰市场的日前优化调度问题。上述文献仅停留在 DER 个体的统一化建模层面, 并未研究如何表征 DERs 集群后的 VPP 聚合模型。

同时, 准确感知其他主体策略突然变化时对 VPP 自身利益的影响趋势也成为一大难题。较多研究聚焦于使用基于模型的运筹学方式, 文献[8-9]提出多种分布鲁棒优化模型来应对多重不确定性对自身优化调度决策的影响。文献[10]提出一种基于鲁棒模型预测控制方法以应对可再生能源的波动性。然而, 上述方法的本质是采用略保守的方式来应对其他主体策略突然变化时对自身收益的影响, 未感知该影响的动力学趋势并实现自趋优决策。

因此, 表征系统动力学过程以获取系统真实状态的自组织演化趋势成为一个新思路。文献[11-13]提出随机动力学的概念。随机动力学将复杂系统的变化表述为“状态变化”, 并将系统的个体状态变化过程数学建模为一个概率驱动的非线性微分方程。其中, 用特定的微分项来表征其他主体策略突然变化带来的状态转换和该转换后涌现新状态的过程。文献[13]利用随机动力学进行选举趋势预测建模, 描绘舆论、信息变化的选举动力学过程。该类方法在电力系统的应用尚处于起步阶段, 在文献[14]中, 笔者采用了一阶非线性微分方程来解决其他主体策略突然变化对自身影响较小的多微网-配电网系统协同问题。然而, 对自身影响较大的多主体系统动力学过程建模问题亟待继续研究。

另一方面, VPP 内 DERs 模型的不准确性对基于模型的传统优化算法也提出了挑战^[15]。具体来说, 基于模型的随机优化和分布鲁棒优化等方法经过一系列改进^[16-17]虽然可以提高不确定性下电力系统调度的决策鲁棒性, 但是该类方法由于内部 DERs 模型的不准确性会使得 VPP 的实际输出与优化算法发布的调度指令总是存在偏差^[18], 从而影响 MVPP 系统的经济安全运行。同时, 上述基于模型的优化算法面对 MVPP 系统复杂的决策变量时存

在收敛速度慢甚至无法收敛的问题, 难以满足其他主体策略突然变化时 VPP 自身所需的较短决策时效。上述问题推动了无模型的多智能体强化学习算法近年来在 VPP 能量管理领域的应用^[19-21]。强化学习算法通过离线训练-在线测试的方式来逐步拟合最优策略, 避免了基于模型的方法可能存在的迭代寻优与病态无解问题, 大幅降低了决策时间。然而, 上述研究只考虑了正常运行状态下的 VPP 或 MVPP 的能量管理问题, 对于 MVPP 系统遭遇内部不确定性影响后的自趋优问题几乎没有涉及。

综上分析, 本文提出一种基于二阶随机动力学的多虚拟电厂自趋优策略, 具体工作包括:

1) 针对 DERs 的异质运行特性, 本文聚焦 DER 可调空间构建了 VPP 资源集群聚合运行模型, 从而有效表征 VPP 的对外特性。

2) 为增强在其他主体策略变化时 VPP 对自身综合收益变化趋势的感知能力, 提出一个二阶随机动力学方程(stochastic dynamic equation, SDE)来捕获 VPP 收益结构的自发演化信息, 从而修正 VPP 的综合收益。并将二阶 SDE 修正的 VPP 综合收益作为奖励函数输入强化学习模型以实现自趋优决策。

3) 针对软动作-评价(soft actor-critic, SAC)算法的训练速度缓慢问题, 本文提出一种基于融合网络的 ISAC 算法, 并设定较小的共享学习率, 使之在不损失训练稳定性的前提下提升学习速度。

1 虚拟电厂聚合运行模型

本文的研究对象是一个如图 1 所示的 MVPP 系统, 各虚拟电厂包含的 DERs 有储能、燃气轮机、电动汽车充电站、可控负荷、可再生能源发电设备。各 DER 将自身运行特性与出力信息上报 VPP 后完成 DER 集群组建, VPP 计算可调空间并参与 MVPP 能量管理。本文假设各 DER 间不一定存在直接线路连接, 均由 VPP 平台通过上层配网进行功率交互且配网拥有足够大的线路容量。

1.1 DER 运行模型

根据异质 DERs 的不同可调控程度, 本文将 DERs 分为不可控 DER 与可控 DER。不可控 DER 为可再生能源发电设备。可控 DER 包括储能、燃气轮机、充电站、可控负荷。下文介绍可控 DER 的运行模型具体构建如下:

1.1.1 可控 DER 出力

可控 DER 出力表示为

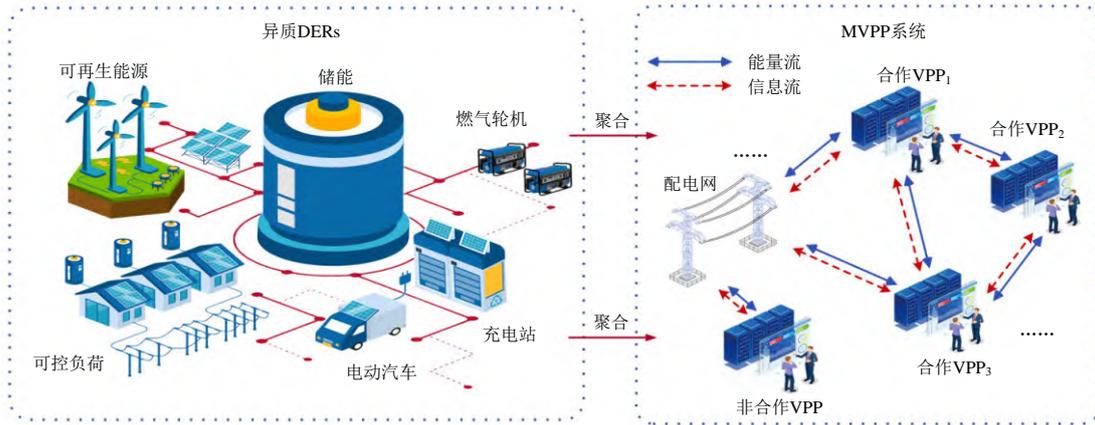


图1 MVPP系统

Fig. 1 MVPP system

$$P_r(k) = \tilde{P}_r(k) + \Delta P_r(k) \quad (1)$$

式中： $P_r(k)$ 为DER在 k 时刻的实际输出功率； $\tilde{P}_r(k)$ 为DER $_r$ 日前计划输出功率； $\Delta P_r(k)$ 为其在日前计划基础上的调节功率。其中， $r \in \{1,2,3,\dots,R\}$ ， $k \in \{1,2,3,\dots,K\}$ 。

1.1.2 可控DER可调空间

可控DER的可调空间由当前时刻的输出功率、能量状态与调节容量共同约束，表示为 $\Delta U_r(k)$ ：

$$\Delta \underline{U}_r(k) \leq \Delta U_r(k) \leq \Delta \bar{U}_r(k) \quad (2)$$

$$\Delta \bar{U}_r(k) = \min\{\bar{P}_r - \tilde{P}_r(k), (\tilde{E}_r(k) - \underline{E}_r)/T_{k \rightarrow k+1}, \bar{W}_r\} \quad (3)$$

$$\Delta \underline{U}_r(k) = \max\{\underline{P}_r - \tilde{P}_r(k), (\tilde{E}_r(k) - \bar{E}_r)/T_{k \rightarrow k+1}, -\bar{W}_r\} \quad (4)$$

式中： $\Delta \bar{U}_r(k)$ 和 $\Delta \underline{U}_r(k)$ 分别为DER $_r$ 可调空间的上、下限； \bar{P}_r 和 \underline{P}_r 分别为DER $_r$ 额定输出功率的上、下限； \bar{E}_r 和 \underline{E}_r 分别为储能型可控DER $_r$ 额定能量容量的上、下限，非储能型可控DER $_r$ 设置为 ∞ ； $\tilde{E}_r(k)$ 为储能型可控DER $_r$ 日前计划中的剩余能量状态，非储能型可控DER $_r$ 设置为0； \bar{W}_r 为DER $_r$ 的允许最大可调节范围； $T_{k \rightarrow k+1}$ 为调度时间间隔，并假设储能型可控DER $_r$ 一旦被调用，持续工作时间即为 $T_{k \rightarrow k+1}$ 。

其中，考虑到储能型DER能量在不同时段存在耦合性质，将剩余能量 $\tilde{E}_r(k)$ 表示为：

$$\tilde{E}_r(k) = \tilde{E}_r(k-1) + T_{k \rightarrow k+1} [\kappa_r^{\text{ch}} \eta_r^{\text{ch}} P_r^{\text{ch}}(k) + \frac{\kappa_r^{\text{dis}} P_r^{\text{dis}}(k)}{\eta_r^{\text{dis}}}] \quad (5)$$

$$\kappa_r^{\text{ch}} + \kappa_r^{\text{dis}} \leq 1 \quad (6)$$

式中： $P_r^{\text{ch}}(k)$ 和 $P_r^{\text{dis}}(k)$ 为DER $_r$ 的充放电功率； η_r^{ch} 和 η_r^{dis} 为DER $_r$ 的充放电效率； κ_r^{ch} 、 κ_r^{dis} 分别为表示充放电状态的0-1变量。

1.1.3 可控DER运行成本

DER运行成本包括增量成本和老化成本。

1) DER增量成本。

日前计划之外的可调空间调控将产生增量运行成本。单个DER增量成本 $c_r^{\text{opc}}(k)$ 表示为：

$$c_r^{\text{opc}}(k) = \varepsilon_r^{\text{opc}} |\Delta P_r(k)| = \varepsilon_r^{\text{opc}} [\Delta P_r^+(k) + \Delta P_r^-(k)] \quad (7)$$

$$\Delta P_r(k) + \Delta P_r^+(k) - \Delta P_r^-(k) = 0 \quad (8)$$

$$\Delta P_r^+(k) \geq 0, \Delta P_r^-(k) \geq 0 \quad (9)$$

式中： $\varepsilon_r^{\text{opc}}$ 为DER $_r$ 的增量运行成本系数； $\Delta P_r^+(k)$ 和 $\Delta P_r^-(k)$ 分别为绝对值形式等价的线性表达。

2) DER老化成本。

频繁的设备调节可能会缩短DER的使用寿命，用老化成本 $c_r^{\text{mil}}(k)$ 来衡量DER的磨损与老化问题：

$$c_r^{\text{mil}}(k) = \varepsilon_r^{\text{mil}} L_r^{\text{mil}} |P_r(k) - P_r(k-1)| \quad (10)$$

式中： $\varepsilon_r^{\text{mil}}$ 为DER $_r$ 调节里程成本系数； L_r^{mil} 为DER $_r$ 平均里程调用率。

1.2 VPP聚合运行模型

规模化DER数量庞大，很难找到一个有效的策略来实现高维管理任务。因此虚拟电厂通过聚合DERs的方式进行统一管理。根据式(1)–(10)给出的DER模型，并考虑DERs的参数特性，构建资源集群架构下的虚拟电厂聚合运行模型，表征虚拟电厂的聚合对外特性。

1.2.1 VPP聚合特性

资源集群架构下VPP出力为各可控DER的总和，可表示为

$$P_i(k) = \tilde{P}_i(k) + \Delta P_i(k) \quad (11)$$

式中： $P_i(k)$ 为VPP $_i$ 在 k 时刻的实际输出功率； $\tilde{P}_i(k)$ 为VPP $_i$ 在 k 时刻的日前计划输出功率； $\Delta P_i(k)$ 为在日前计划外的调节功率，其中 $i \in \{1,2,3,\dots,I\}$ 。

针对单个可控DER $_r$ ，可调空间变量集合可表

示为 $\delta_r = \{\tilde{P}_r(k), \bar{P}_r, \underline{P}_r, \tilde{E}_r(k), \bar{E}_r, \underline{E}_r, \bar{W}_r\}$ 。VPP 的可调空间应为所含所有 DERs 的可调空间总和，可得 VPP_i 可调空间变量集合 δ_i 。

$$\delta_i = \sum_{r \in \Theta_i} \delta_r = \{\tilde{P}_i(k), \bar{P}_i, \underline{P}_i, \tilde{E}_i(k), \bar{E}_i, \underline{E}_i, \bar{W}_i\} \quad (12)$$

式中 Θ_i 为 VPP_i 中的可控 DER 集合。

进一步，单个可控 DER_r 的特性参数集合可表示为 $\vartheta_r = \{\eta_r^{\text{ch}}, \eta_r^{\text{dis}}, \varepsilon_r^{\text{ope}}, \varepsilon_r^{\text{mil}}, L_r^{\text{mil}}\}$ 。VPP 的特性参数应为所含所有 DERs 的特性参数加权和，本文参考文献[22]方法获得各参数权重系数，得到 VPP_i 近似特性参数集合 ϑ_i 。

$$\vartheta_i = \sum_{r \in \Theta_i} w_r \vartheta_r = \{\eta_i^{\text{ch}}, \eta_i^{\text{dis}}, \varepsilon_i^{\text{ope}}, \varepsilon_i^{\text{mil}}, L_i^{\text{mil}}\} \quad (13)$$

式中 w_r 为 DER_r 特性参数的权重系数，满足 $\sum_{r \in \Theta_i} w_r = 1$ 。

1.2.2 VPP 可调空间

由 1.2.1 节所得的聚合特性，VPP_i 的可调空间可表示为 $\Delta U_i(k)$ ，如下：

$$\Delta \underline{U}_i(k) \leq \Delta U_i(k) \leq \Delta \bar{U}_i(k) \quad (14)$$

$$\Delta \bar{U}_i(k) = \min\{\bar{P}_i - \tilde{P}_i(k), (\tilde{E}_i(k) - \underline{E}_i)/T_{k \rightarrow k+1}, \bar{W}_i\} \quad (15)$$

$$\Delta \underline{U}_i(k) = \max\{\underline{P}_i - \tilde{P}_i(k), (\tilde{E}_i(k) - \bar{E}_i)/T_{k \rightarrow k+1}, -\bar{W}_i\} \quad (16)$$

式中： $\Delta \bar{U}_i(k)$ 和 $\Delta \underline{U}_i(k)$ 分别为 VPP_i 可调空间的上、下限； \bar{P}_i 和 \underline{P}_i 分别为 VPP_i 额定输出功率的上、下限； \bar{E}_i 和 \underline{E}_i 分别为 VPP_i 额定能量容量的上、下限； $\tilde{E}_i(k)$ 为 VPP_i 日前计划中的剩余能量状态； \bar{W}_i 为 VPP_i 的允许最大可调节范围。

其中，剩余能量 $\tilde{E}_i(k)$ 表示为：

$$\tilde{E}_i(k) = \tilde{E}_i(k-1) + T_{k \rightarrow k+1} [\kappa_i^{\text{ch}} \eta_i^{\text{ch}} P_i^{\text{ch}}(k) + \frac{\kappa_i^{\text{dis}} P_i^{\text{dis}}(k)}{\eta_i^{\text{dis}}}] \quad (17)$$

$$\kappa_i^{\text{ch}} + \kappa_i^{\text{dis}} \leq 1 \quad (18)$$

式中： $P_i^{\text{ch}}(k)$ 和 $P_i^{\text{dis}}(k)$ 为 VPP_i 的充放电功率； η_i^{ch} 和 η_i^{dis} 为 VPP_i 的充放电效率； κ_i^{ch} 和 κ_i^{dis} 分别为表示充放电状态的 0-1 变量。

1.2.3 VPP 运行成本

聚合后的 VPP_i 运行成本由 $C_i^{\text{vpp}}(k)$ 由增量成本 $C_i^{\text{ope}}(k)$ 与老化成本 $C_i^{\text{mil}}(k)$ 组成，近似表示为：

$$C_i^{\text{ope}}(k) = \varepsilon_i^{\text{ope}} |\Delta P_i(k)| \quad (19)$$

$$C_i^{\text{mil}}(k) = \varepsilon_i^{\text{mil}} L_i^{\text{mil}} |P_i(k) - P_i(k-1)| \quad (20)$$

式中： $\varepsilon_i^{\text{ope}}$ 为 VPP_i 的增量运行成本系数； $\varepsilon_i^{\text{mil}}$ 为 VPP_i 调节里程成本系数； L_i^{mil} 为 VPP_i 平均里程调用率。

2 基于随机图理论的 MVPP 互动网络

2.1 随机图架构下的互动网络建模

MVPP 的能量互动架构是一个主体间存在随机合作关系的点对点网络，符合广义随机图理论的数学定义^[23]。也就是说，VPP 间的随机合作关系可以被视作图内节点间的随机连接过程。在每一个时刻，所有 VPP 需要选择互动身份，身份包括合作者种群和非合作者种群。合作者种群代表参与能量互动；非合作者种群代表独立运行，不参与能量互动。

1) MVPP 状态随机图的数学描述：多虚拟电厂状态随机图用 $\mathbf{G}(k)=[\mathbf{X}(k), \mathbf{L}(k)]$ 来描述。 $\mathbf{X}(k) \in \mathbb{R}^{N \times N}$ 是 MVPP 系统的邻接合作矩阵，用来存储各虚拟电厂的随机合作概率信息。此外， $\mathbf{L}(k) \in \mathbb{R}^{N \times 1}$ 是 MVPP 系统的策略信息矩阵，用来存储各虚拟电厂的气象及策略信息。

2) 邻接合作矩阵的数学描述：因为本文邻接关系的虚拟属性，采用随机合作信息表示邻接矩阵 $\mathbf{X}(k)$ ，即 $\mathbf{X}(k) = (x_1(k), x_2(k), \dots, x_N(k))^T$ ，其中 $x_i(k) \in \{0, 1\}$ 表示 k 时刻 VPP_i 为合作者的概率。如 2.1 节中所述，本文引入随机图论中的平均度 $d(k)$ 来表征系统的已合作程度。平均度 $d(k)$ 为

$$d(k) = (I-1)\bar{x}(k) = \frac{(I-1)}{I} \sum_{i \in I} x_i(k) \quad (21)$$

3) 策略信息矩阵的数学描述：策略信息矩阵 $\mathbf{L}_i(k) = [F_i^{\text{VPP}}(k), \mathbf{D}_i(k), \mathbf{M}_i^{\text{WT}}(k), \mathbf{M}_i^{\text{PV}}(k)]$ 存储了当前系统各虚拟电厂的策略信息。包括收益状态 $F_i^{\text{VPP}}(k)$ ，表示当前虚拟电厂收益；控制变量 $\mathbf{D}_i(k) = [P_i^{\text{VPP}}(k), P_i^{\text{ET}}(k)]$ ， $P_i^{\text{VPP}}(k)$ 表示 VPP_i 自身设备调节功率， $P_i^{\text{ET}}(k)$ 表示与其他主体的能量互动功率；设备参数及气象信息 $\mathbf{M}_i^{\text{WT}}(k)$ 、 $\mathbf{M}_i^{\text{PV}}(k)$ 。

2.2 合作-非合作群组成本函数

针对两类群组不同的收益结构，本文提出一种基于个体贡献的补贴机制^[24]。合作者因为参与功率交互以提升新能源消纳率可以获得较多的补贴收益。非合作者虽然未参与能量交易，但是可以作为系统的关键备用以获得较少的补贴。同时，不同决策时刻的身份转变也将造成决策平台的额外决策成本，因此个体将支付身份转变成本。

2.2.1 合作者 VPP 的综合成本函数

1) 运行成本函数。

合作者虚拟电厂运行成本函数 $C_i^{\text{vpp}}(k)$ 由虚拟电厂增量成本 $C_i^{\text{ope}}(k)$ 与老化成本 $C_i^{\text{mil}}(k)$ 组成，表

示为

$$C_i^{\text{vpp}}(k) = C_i^{\text{opc}}(k) + C_i^{\text{mil}}(k) \quad (22)$$

2) 能量交互成本函数。

虚拟电厂能量交互成本函数 $C_i^{\text{et}}(k)$ 由虚拟电厂与其他虚拟电厂交互成本 $C_i^{\text{et-vpp}}(k)$ 和与上层配电网交互成本 $C_i^{\text{et-main}}(k)$ 组成, 被表示为

$$C_i^{\text{et}}(k) = C_i^{\text{et-vpp}}(k) + C_i^{\text{et-dn}}(k) \quad (23)$$

与其他虚拟电厂交互成本 $C_i^{\text{et-vpp}}(k)$ 主要考虑向其他虚拟电厂购买或售出的交互成本以及需向配网支付的网络占用费, $C_i^{\text{et-vpp}}(k)$ 为:

$$C_i^{\text{et-vpp}}(k) = P_i^{\text{pur}}(k)c_i^{\text{pur}}(k) - P_i^{\text{sell}}(k)c_i^{\text{sell}}(k) + P_i^{\text{et-vpp}}(k)c^{\text{tra}}(k) \quad (24)$$

$$P_i^{\text{et-vpp}}(k) = P_i^{\text{pur}}(k) + P_i^{\text{sell}}(k) \quad (25)$$

式中: $P_i^{\text{pur}}(k)$ 和 $P_i^{\text{sell}}(k)$ 分别为购买与售出的功率; $P_i^{\text{et-vpp}}(k)$ 为总交互功率; $c_i^{\text{pur}}(k)$ 和 $c_i^{\text{sell}}(k)$ 分别为购买与售出的单位电价, 在本文中设为定值, 为确保交易隐私仅交易双方知晓; $c^{\text{tra}}(k)$ 为需向配网支付的网络占用单价。

其中, 本文设定与上层配网交互功率的优先级低于与其他虚拟电厂交互, $C_i^{\text{et-dn}}(k)$ 为

$$C_i^{\text{et-dn}}(k) = [P_i^{\text{load}}(k) + P_i^{\text{pur}}(k) - P_i^{\text{sell}}(k)]c_i^{\text{dn}}(k) \quad (26)$$

式中: $P_i^{\text{load}}(k)$ 为该虚拟电厂的能量服务计划; $c_i^{\text{dn}}(k)$ 为向配网购电的单位电价。

3) 身份转变成本函数。

不按照日前计划的突然身份转变也将造成决策平台的额外决策成本, 因此个体将支付身份转变成本 $C_i^{\text{cha}}(k)$, 表示为

$$C_i^{\text{cha}}(k) = |\alpha_i(k-1) - \alpha_i(k)|c^{\text{nc}} \quad (27)$$

式中: $\alpha_i(k)$ 为身份校验参数, 当虚拟电厂为合作者时 $\alpha_i(k)=1$, 反之, $\alpha_i(k)=0$; c^{nc} 为合作者转变为非合作者所需支付的额外决策费用。

4) 互动补贴收益函数。

参与合作机制的虚拟电厂可以获得补贴收益以促进合作。针对现有方法追求个体利益而忽略系统合作资源的可持续性问题, 本文引入动态协同因子以量化系统合作资源的可持续性。补贴收益函数 $T_i^{\text{sub}}(k)$ 表示为:

$$T_i^{\text{sub}}(k) = \frac{c^{\text{sub-c}}}{I} \sum_{i \in I} \varphi_i(k) \alpha_i(k) P_i^{\text{et-vpp}} \quad (28)$$

$$\varphi_i(k) = \varphi_i(k-1) + \log_{\omega}[d(k)] - \log_{\omega}[d(k-1)] \quad (29)$$

式中: $\varphi_i(k) \geq 1$ 为系统动态协同因子, 用来量化 MVPP 系统内资源的已合作程度进而影响利益分配结果; ω 为定值波动系数, 用来表征系统协同因子的波动规模; $c^{\text{sub-c}}$ 为定值价格系数, 表示系统合作者补贴收益单价。

5) 合作者综合成本函数。

综合上述, 合作者虚拟电厂的综合成本函数 F_i^{CVPP} 表示为

$$F_i^{\text{CVPP}} = \sum_{k=1}^K C_i^{\text{vpp}}(k) + C_i^{\text{et}}(k) + C_i^{\text{cha}}(k) - T_i^{\text{sub}}(k) \quad (30)$$

2.2.2 非合作者 VPP 的综合成本函数

1) 运行成本函数。

非合作者虚拟电厂的运行成本函数同式(22)。

2) 身份转变成本函数。

非合作者虚拟电厂的运行成本函数 $C_i^{\text{cha}}(k)$ 表示为

$$C_i^{\text{cha}}(k) = |\alpha_i(k-1) - \alpha_i(k)|c^{\text{nc-c}} \quad (31)$$

式中 $c^{\text{nc-c}}$ 为非合作者转变为合作者所需支付的额外决策费用。

3) 备用补贴收益函数。

非合作者可以选择作为系统的关键备用以获得较少的补贴, 且补贴来源于合作者群组。补贴收益函数 $T_i^{\text{sub}}(k)$ 表示为

$$T_i^{\text{sub}}(k) = \frac{c^{\text{sub-nc}}}{I} \sum_{n \in I^{\text{CVPP}}(k)} T_n^{\text{sub}}(k) \quad (32)$$

式中: $I^{\text{CVPP}}(k)$ 为合作者虚拟电厂集合; $c^{\text{sub-nc}}$ 为定值价格系数, 表示系统非合作者补贴收益单价。

4) 非合作者综合成本函数。

综合上述, 非合作者虚拟电厂的综合成本函数 F_i^{NCVPP} 表示为

$$F_i^{\text{NCVPP}} = \sum_{k=1}^K C_i^{\text{vpp}}(k) + C_i^{\text{cha}}(k) - T_i^{\text{sub}}(k) \quad (33)$$

3 基于随机动力学的 MVPP 自趋优模型

为了增强 VPP 应对其他主体策略变化时的自趋优性能, 本文将系统真实收益的求取问题数学表征为一个概率特性驱动的随机动力学模型。具体地, 构建一个随机动力学方程, 并用方程中特定的微分项表征其他主体策略突然变化带来的状态转换和该转换后涌现新状态的过程。最后设定边界条件求解方程, 获取系统状态变化的数值趋势。本问设定 VPP 综合成本为系统的“状态”, 综合成本的

变化为系统的“状态变化”。

3.1 二阶随机动力学方程推导

3.1.1 数学准备与定义

1) 状态与序参量。

定义 1：定义 VPP_i 状态为综合成本函数 $F_i^{VPP}(k)$ ；序参量为虚拟电厂状态上升幅值 $f_i^{inc}(k)$ 和下降幅值 $f_i^{dec}(k)$ 。

在自组织理论中，序参量是描述状态变化趋势的参数^[12]。由式(30)和(33)易知，上升幅值 $f_i^{inc}(k)$ 由能量交互收益和补贴收益所影响，下降幅值 $f_i^{dec}(k)$ 由运行成本和身份转变成成本所影响。

假设 1：将相邻时刻的中间时间 $T_{k \rightarrow k+1}$ 等分成 h 个很小的时间间隔 κ ，即 $T_{k \rightarrow k+1} = h\kappa$ ($h=1, 2, 3, \dots$)。假设 $f_i^{inc}(k)$ 和 $f_i^{dec}(k)$ 在相邻时刻内的变化速度是均匀的，并将 κ 内的 $f_i^{inc}(k)$ 和 $f_i^{dec}(k)$ 的变化记作 Δf_i^{inc} 和 Δf_i^{dec} 。

2) 状态变化。

定义 2：定义在 $T_{k \rightarrow k+1}$ 内的特定离散时间点 n ，虚拟电厂状态变化幅值为 f_i 的概率为 $P(f_i, n)$ ，其中，($1 \leq n \leq h$)。

每个决策时刻个体状态均存在发生变化的概率，且每一时刻状态会发生变化的概率仅与上一时刻状态变化的概率相关。因此，可以给出以下引理。

引理 1：根据上述定义与假设，可以给出状态变化方程如式(34)所示，推导过程见附录 A 式(A1)、(A2)。

$$P(f_i, n) = P(f_i - \Delta f_i^{inc}, n-1) + P(f_i + \Delta f_i^{dec}, n-1) - P(f_i, n-1) \quad (34)$$

$P(f_i - \Delta f_i^{inc}, n-1)$ 和 $P(f_i + \Delta f_i^{dec}, n-1)$ 将系统策略变化引发的状态变化融入到状态变化方程中，并且通过对上一时刻状态的迭代计算使上一时刻的状态自发成为下一时刻状态的来源。

3.1.2 二阶随机动力学方程

引理 2：同理， $n+1$ 时刻的状态变化方程可以表示为式(35)：

$$P(f_i, n+1) = P(f_i - \Delta f_i^{inc}, n) + P(f_i + \Delta f_i^{dec}, n) - P(f_i, n) \quad (35)$$

将式(34)代入式(35)，并进行依次迭代。对于任意时刻 $n+\tau$ ，虚拟电厂 i 状态变化幅值为 f_i 的概率 $P(f_i, n+\tau)$ 可以被推导为式(36)：

$$P(f_i, n+\tau) = \begin{cases} 0, & \text{当 } \tau - \gamma_1 - \gamma_2 \leq 0 \\ \sum_{\gamma_1, \gamma_2=0}^{\tau} \frac{(-1)^{(\tau-\gamma_1-\gamma_2)} \tau! P(f_i - \gamma_1 \Delta f_i^{inc} + \gamma_2 \Delta f_i^{dec}, n)}{\gamma_1! \gamma_2! (\tau - \gamma_1 - \gamma_2)!}, & (36) \\ \text{其他} \end{cases}$$

然后，对 $P(f_i, n+\tau)$ 进行泰勒级数展开。如上文所述，自组织理论认为二阶相变为自组织过程的对应物^[12-13]。本文为了保证泰勒分解只在 f 点附近，对展开的泰勒级数保留到二阶。二阶 SDE 表示如下：

$$\frac{dP(f_i, t)}{dt} = \mu_{1,i}(k) \frac{d^2 P(f_i, t)}{df_i^2} - \mu_{2,i}(k) \frac{dP(f_i, t)}{df_i} - \mu_{3,i}(k) \frac{d^2 P(f_i, t)}{dt^2} \quad (37)$$

$$\begin{cases} \mu_{1,i}(k) = [f_i^{inc}(k)^2 - f_i^{inc}(k)f_i^{dec}(k) + f_i^{dec}(k)^2] \\ \mu_{2,i}(k) = [f_i^{inc}(k) - f_i^{dec}(k)] \\ \mu_{3,i}(k) = 1 \end{cases} \quad (38)$$

式中： $dP(f_i, t)/df_i$ 为正常情况下状态变化的过程； $[d^2 P(f_i, t)]/(df_i^2)$ 为随机策略改变存在下状态变化的过程； $dP(f_i, t)/dt$ 为状态变化的速度； $d^2 P(f_i, t)/dt^2$ 为随机策略改变造成的状态变化后涌现新状态的自组织过程。 $\mu_{1,i}(k)$ 、 $\mu_{2,i}(k)$ 和 $\mu_{3,i}(k)$ 为随机动力学方程的动态系数。

3.1.3 边界条件

为了求解微分方程，本文对(37)设定边界条件，见附录 B 式(B1)—(B3)。

4 MVPP 系统自趋优能量管理强化学习决策方法

针对 MVPP 系统能量管理中的离散时间特征与主体策略变化问题，本文采用基于 ISAC 强化学习决策框架。首先，基于蒙特卡洛法构造 MVPP 系统正常运行数据集对 ISAC 网络进行离线训练；然后，利用随机动力学方程感知其他主体策略变化下的 VPP 综合收益变化趋势，并依次修正 VPP 真实收益值；最后将修正收益作为 ISAC 决策架构的奖励进行策略求解。

4.1 多智能体动作空间、状态空间、奖励函数

定义 3：定义从环境中学习到的信息为状态 S ；前一状态下的执行情况为动作 A ；根据状态决定的反应为策略 π ： $S \rightarrow A$ ；环境对特定动作的评价反馈为奖励 R 。

1) 智能体与环境：设定 VPP_i 为智能体，MVPP

系统为环境。其中 $i \in I = \{1, 2, 3, \dots, I\}$ 。

2) 动作空间: $a_k = \{a_{i,k} | i \in I\}, \forall a_k \in \mathbf{A}$ 为 MVPP 系统在 k 时刻的联合动作空间; 其中 $a_{i,k} = [D_i(k)]$ 为 k 时刻 VPP_{*i*} 的动作空间, 包括 VPP_{*i*} 自身设备调节功率和与其他合作 VPP 的能量互动功率。

3) 状态空间: $s_k = \{s_{i,k} | i \in I\}, \forall s_k \in \mathbf{S}$ 为 MVPP 系统在 k 时刻的联合状态空间; 其中 $s_{i,k} = [\Delta U_i(k), \mathbf{M}_i^{\text{WT}}(k), \mathbf{M}_i^{\text{PV}}(k), x_i(k)]$ 为 k 时刻 VPP_{*i*} 的状态空间, 包括可调空间、气象信息以及身份信息。

4) 奖励函数: $r_k = \{\sum_{i \in I} r_{i,k}, \forall r_k \in \mathbf{R}\}$ 为 k 时刻随机动力学方程得到的 MVPP 系统修正综合收益; 其中, $r_{i,k} = [-\tilde{F}_i^{\text{VPP}}(k)]$ 。

$$-\tilde{F}_i^{\text{VPP}}(k) = -F_i^{\text{VPP}}(k) + \Delta F_i^{\text{VPP}}(k) \quad (39)$$

式中: $-\tilde{F}_i^{\text{VPP}}(k)$ 为 VPP_{*i*} 由随机动力学方程求解得到的修正综合收益; $\Delta F_i^{\text{VPP}}(k)$ 为修正量。

需要注意的是, 区别于 3.2 节所述的虚拟电厂状态, 此处 \mathbf{S} 为深度强化学习决策过程中的状态空间。

4.2 多智能体 ISAC 算法

4.2.1 SAC 算法原理

在离散空间中, 熵被用来衡量策略的随机性。当每一种随机事件可能发生的概率相等, 即随机事件服从均匀分布时熵最大。SAC 算法面对随机改变时融入了最大熵机理, 帮助智能体尽可能地随机选择动作空间, 增强其对任何有利动作的探索性以防陷入局部最优^[25]。

SAC 算法包含一个动作网络和一组孪生的评价 Q 网络, 其结构图如附图 A1 所示。从状态空间 \mathbf{S} 映射到动作空间 \mathbf{A} 上的最大熵策略函数累计奖励可以表示为

$$\pi^* = \arg \max_{\pi} V^{\pi}(s) = \mathbb{E}_{a_k \sim \pi(\cdot | s_k)} \sum_{\tau=0}^{\infty} v^{\tau} [r_k(s_k, a_k) - \alpha \log \pi(a_k | s_k)] \quad (40)$$

式中: a_k 和 s_k 分别为 k 时刻的动作和状态空间; $\pi(a_k | s_k)$ 为从当前状态空间 s_k 映射到动作空间 a_k 上的一个分布的策略; $v^k \in (0, 1)$ 为折现系数; α 为熵正则化系数, 用来决定熵与奖励的权重占比。

考虑熵自动调整^[26]后, 熵正则化系数 α 的梯度如式(41)所示:

$$J(\alpha) = \mathbb{E}_{a_k \sim \pi_k} [-\alpha \log \pi_k(a_k | s_k) - \alpha \bar{H}] \quad (41)$$

式中 \bar{H} 表示期望的最小期望熵。

将评价 Q 网络用 θ 进行参数化表示, 从而将评

价 Q 网络目标函数表示为 $J_Q(\theta)$:

$$J_Q(\theta) = \mathbb{E}_{(s_k, a_k) \sim D} \left\{ \frac{1}{2} [Q_{\theta}(s_k, a_k)(r(s_k, a_k) + v^k \mathbb{E}_{s_{k+1} \sim p} V_{\bar{\theta}}(s_{k+1}))]^2 \right\} \quad (42)$$

式中: D 为采样状态和动作的分布重放缓冲区; $\bar{\theta}$ 为目标 Q 网络, 数值上等于 Q 网络的指数移动平均值。

将动作网络 $\pi_{\phi}(a_k | s_k)$ 用 ϕ 进行参数化表示, 从而将动作网络目标函数表示为 $J_{\pi}(\phi)$:

$$J_{\pi}(\phi) = \mathbb{E}_{s_k \sim D} \{ \mathbb{E}_{a_k \sim \pi_{\phi}} [\alpha \log \pi_{\phi}(a_k | s_k) - Q_{\theta}(s_k, a_k)] \} \quad (43)$$

各智能体网络按梯度下降方向更新参数, 各网络对应的目标函数梯度如式(44)—(46)所示:

$$\hat{V} J(\alpha) = -\log \pi_{\phi}(a_k | s_k) - \bar{H} \quad (44)$$

$$\hat{V}_{\phi} J_{\pi}(\phi) = \nabla_{\phi} V_{\phi}(s_k) [V_{\phi}(s_k) - Q_{\theta}(s_k, a_k) + \log \pi_{\phi}(a_k | s_k)] \quad (45)$$

$$\hat{V}_{\theta} J_Q(\theta) = \nabla_{\theta} Q_{\theta}(a_k, s_k) [Q_{\theta}(s_k, a_k) - r_k(s_k, a_k) - v V_{\bar{\theta}}(s_{k+1})] \quad (46)$$

4.2.2 多智能体 ISAC 算法设计

在随机改变的影响下, 包括 SAC 在内的强化学习算法通常都存在学习缓慢的问题, 而对网络盲目的合并或者剪枝会导致算法失去稳定性。为此, 本文提出一种融合软动作-评价(integrated soft actor-critic, ISAC)算法, 改进后的结构图如图 2 所示。

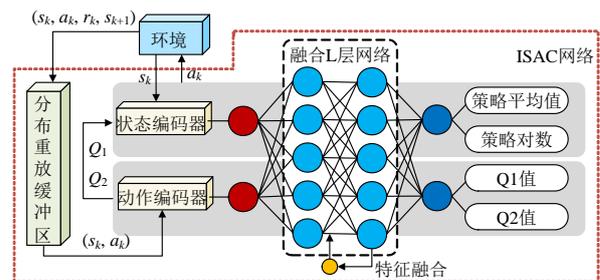


图 2 ISAC 算法决策框架

Fig. 2 ISAC algorithm decision-making framework

在 SAC 中, 策略网络和 2 个孪生评价 Q 网络完全独立, 但是它们拥有相似的网络结构。本文将 2 个 Q 网络合并, 再与策略网络共享输入和训练网络。其中, 张量 M 表示输入或输出的维度, 具体改进方法为:

1) 将原始 SAC 中的输入输出转换为可以直接相加的编码器。对输入输出层的模块化处理便于网络组合时进行串联或加法运算。在不损失性能的情况下, 加法运算可以提高学习速度。

2) 合并原始 SAC 中 3 个网络的 L 层网络, 并仅在使用时激活。图 2 中的融合 L 层网络由动作网络与 2 个评价 Q 网络的主干合并而来。在实际训练时, 动作网络与评价网络共享状态编码器与融合 L 层网络, 从而降低网络复杂度。

3) 在合并后的网络中加入特征联合。本文以张量叠加作为连接方式, 将各网络提取的特征进行维度融合, 增强网络的特征提取和学习能力, 从而避免由于合并网络导致的训练不收敛。

如图 2 所示, 首先智能体 i 观测所在 VPP 的状态空间 $s_{i,k}$, 并根据策略函数 $\pi(a_{i,k}|s_{i,k})$ 生成动作空间 $a_{i,k}$ 。接着, 依次完成所有智能体动作后, 系统根据联合动作空间 a_k 向各智能体反馈奖励值。然后, 系统过渡到下一状态 s_{k+1} , 采集所有智能体当前状态空间、动作空间、奖励存储于分布重放缓冲区。重复上述交互过程直至所有智能体不再改变策略函数, 此时多智能体动作策略函数即为系统能量管理策略。

4.3 求解流程

本文构建一个随机动力学方程-多智能体 ISAC 算法的两阶段求解框架。基于随机动力学方程的修正奖励流程如附图 E1 所示。多智能体 ISAC 算法流程如附图 E2 所示, 其中学习率均设置为原始全局学习率的 0.5 倍以适应网络改进。多智能体与环境进行交互, 并依据附图 E1 获得修正奖励。

5 算例分析

5.1 算例设置

为了测试多虚拟电厂自组织能量管理优化的结果, 算例设置 3 个 VPP 组成 MVPP 系统, 分别记为 VPP₁—VPP₃。算例中各 VPP 中 DER 数量如附表 C1 所示。各 VPP 所含 DERs 对应的相关参数由蒙特卡洛抽样方法根据附表 C2—C6 概率分布获得。价格系数如附表 C8 所示。实时调度的时间间隔设置为 15 min。采用蒙特卡洛法构造 MVPP 系统连续 500 d 的日常运行数据, 并随机抽样主体策略变化数据添加到运行数据中循环生成多智能体训练集。随机选取 100 d 作为测试集。

本文所提出算法在 PyCharm 平台上以 Python 3.8 实现, 采用 Numpy 库求解非线性微分方程, 并使用 PyTorch 框架进行训练。ISAC 算法中超参数设定如附表 C7 所示。硬件配置为 AMD R93950X 处理器, RTX4090 显卡, 内存 32 GB。

5.2 训练过程分析

为保证网络收敛, 设置 ISAC 训练步数为 5 000 步。附图 D1 为 ISAC 训练过程中的损失函数变化曲线。图 3 为 ISAC 训练过程中的奖励函数变化曲线。各智能体的奖励函数值均在 1 500 步后基本收敛, VPP₁、VPP₂、VPP₃ 的奖励函数分别收敛至 2.35、8.65、10.05 附近。由于 VPP₃ 中含 DER 数量最少, 所获的综合收益相对较小, 因此奖励函数最终收敛值小于 VPP₁ 和 VPP₂。

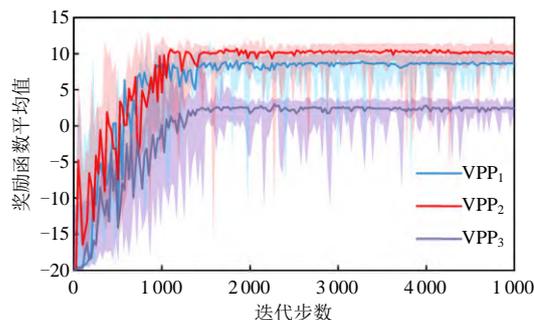


图 3 多智能体 ISAC 算法训练奖励

Fig. 3 Training rewards of multi-agent ISAC algorithm

为验证 ISAC 算法能在不失稳定性的情况下提高训练速度, 本文采用原始 SAC 算法与之对比, 训练过程中奖励函数变化曲线如图 4 所示。ISAC 算法在 1 000 步附近开始收敛, 1 200 步后基本稳定, 而 SAC 算法在 2 000 步附近才趋于稳定, 表明 ISAC 算法拥有更快的收敛速度。ISAC 算法与 SAC 算法的奖励函数值均稳定于 7 左右, 且阴影部分面积大小与变化趋势近似相同, 表明网络合并后的 ISAC 算法依然具有较好的稳定性。

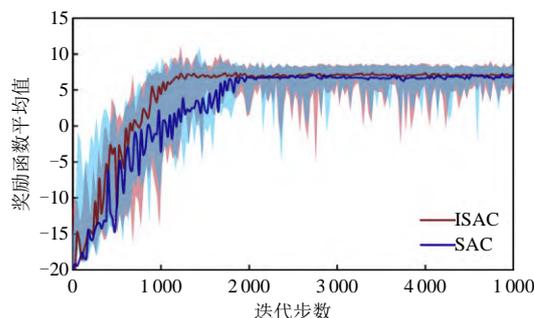


图 4 多智能体 ISAC 与 SAC 训练过程对比

Fig. 4 Comparison of training processes between multi-agent ISAC and SAC

为了进一步验证 ISAC 算法的稳定性, 本文在不同学习率倍数下采用 SAC 算法与之对比。SAC 与 ISAC 算法收敛后, 计算损失函数最大值和最小值之差的平均值, 求以其比值表示 ISAC 稳定性。测试过程中数据记录如附表 D2 所示。在收敛范围

内调小学习率后, SAC 收敛步数从 1600 增加至 2000, ISAC 收敛步数从 1100 增加至 1500, ISAC 稳定性提升 15.9%, 与 SAC 算法仅相差 3.5%。进一步证明恰当的学习率下, ISAC 算法能在不失稳定性的情况下提高训练速度。

5.3 能量管理结果

5.3.1 正常运行下的能量管理结果

正常运行时 VPP₁—VPP₃ 在一阶 SDE 和二阶 SDE 驱动下的能量管理结果如图 5 所示。考虑到自身 DERs 数量及容量的限制, VPP₁ 在整个调度周期均为合作者身份。00:00—09:00 和 19:00—22:00 风电出力较强, 因此在完成能量服务后可以进行设备充电或出售给其他 VPP。09:00—19:00 时由于光伏设备的缺失, VPP₁ 需要通过购入能量或设备放电来完成能量服务。然而, 频繁的调节会带来设备的老化与高昂的增量运行费用, 因此经济的方式是通过购买其他主体的多余新能源发电能量。由图 5(a)可

知, 一阶 SDE 驱动下的设备调节频率比二阶 SDE 驱动下的设备调节频率高。

另一方面, 由于自身 DERs 数量及容量较大, VPP₂ 在整个调度周期均选择非合作者身份。值得注意的是, 由图 5(b)可知, 一阶 SDE 驱动下的能量管理策略会选择向配网购买较少高价电, 而二阶 SDE 驱动下的能量管理策略可以将这部分购电任务更好地分配给可控 DERs 以保证收益的提升。同时, 相较于 VPP₂, VPP₃ 的可再生能源资源数量与容量较小, 但种类较全。正常运行下, VPP₃ 在整个调度周期均为合作者身份, 与 VPP₁ 交互功率以促进各自可再生能源消纳。VPP₃ 在 00:00—09:00 和 19:00—22:00 风电出力较弱, 需向 VPP₁ 购电或使用可控设备放电。区别于 VPP₁, VPP₃ 具有较大的光伏容量, 可以在 09:00—15:00 向 VPP₁ 售电或使用可控设备充电。相同地, 由图 5(c)可知, 一阶 SDE 驱动下的设备调节频率比二阶 SDE 驱动下的设备调节频率高。

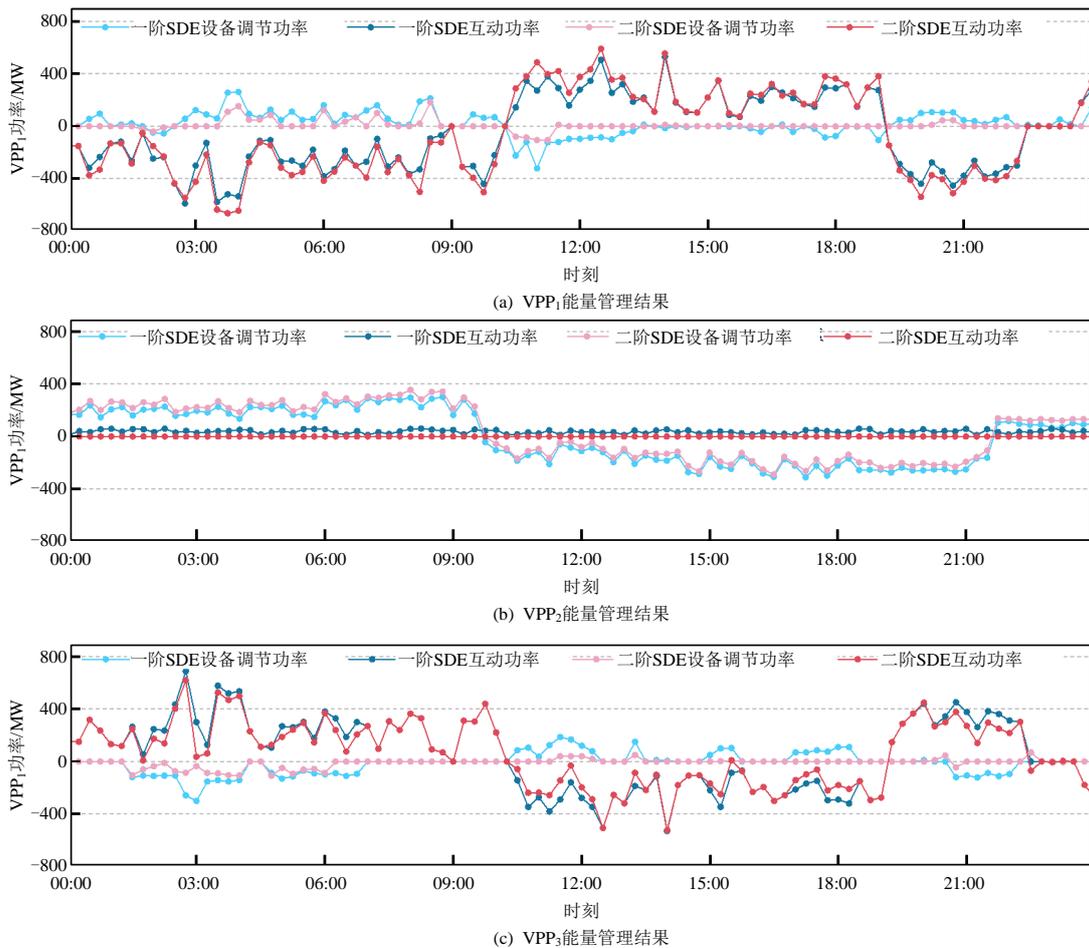


图 5 正常运行下 MVPP 系统能量管理结果

Fig. 5 Energy management results of MVPP system under normal operation

5.3.2 其他主体策略变化时的能量管理结果

为模拟其他主体策略变化时的 MVPP 能量管

理结果, 本文设计主体策略变化为: 在 20:15—21:30 出现未预测到的风力加强从而风电大发, 导致 VPP₃

在个体利益驱动下决策成为非合作者。VPP₃ 策略突然变化时 VPP₁—VPP₃ 在一阶 SDE 和二阶 SDE 驱动下的能量管理结果如图 6 所示。由于风电大发，VPP₃ 仅需少量调节自身设备就可以完成能量服务。由图 6(c)可知，VPP₃ 在 20:15—21:30 成为非合作者，且一阶和二阶 SDE 驱动下的能量管理结果与图 6(b)中 VPP₂ 相似。

然而，VPP₃ 的策略突然转变会对 VPP₁ 和 VPP₂ 的能量管理决策造成较大影响。从一阶 SDE 驱动的能量管理结果来看，VPP₂ 在转变发生后依然保持了非合作者身份，作为合作者的 VPP₁ 在该转变下失去了合作对象，只能通过可控设备充电来消纳多余的可再生能源能量。因此，这是一种被动的应对方式，即通过价格高昂的设备调节方式来应对。然而，从二阶 SDE 驱动的能量管理结果来看，VPP₂ 在

20:15—21:30 中发现“商机”，自发地完成身份转变，与 VPP₁ 自趋地形成了新的合作者联盟，以合作者的身份购买了 VPP₁ 所剩余的全部可再生能源能量，从而减小了在这段时间内的设备调节频率。这是因为二阶 SDE 对其他主体策略转变后收益变化趋势与新收益状态涌现概率的精准感知，使多智能体在训练过程中探索了可能存在更高奖励的协同决策方案。所以对于 VPP₃ 策略突然变化的真实测试场景，自趋性地动态联盟或解约将成为一个提升自治能力的更优策略。

21:30 后风力恢复正常，VPP₃ 没有恢复合作者身份，互动功率为 0，通过 20:15—21:30 设备存储的多余能量完成了能量服务任务。VPP₂ 恢复为非合作者身份，VPP₁ 通过调节可控设备完成能量服务任务。

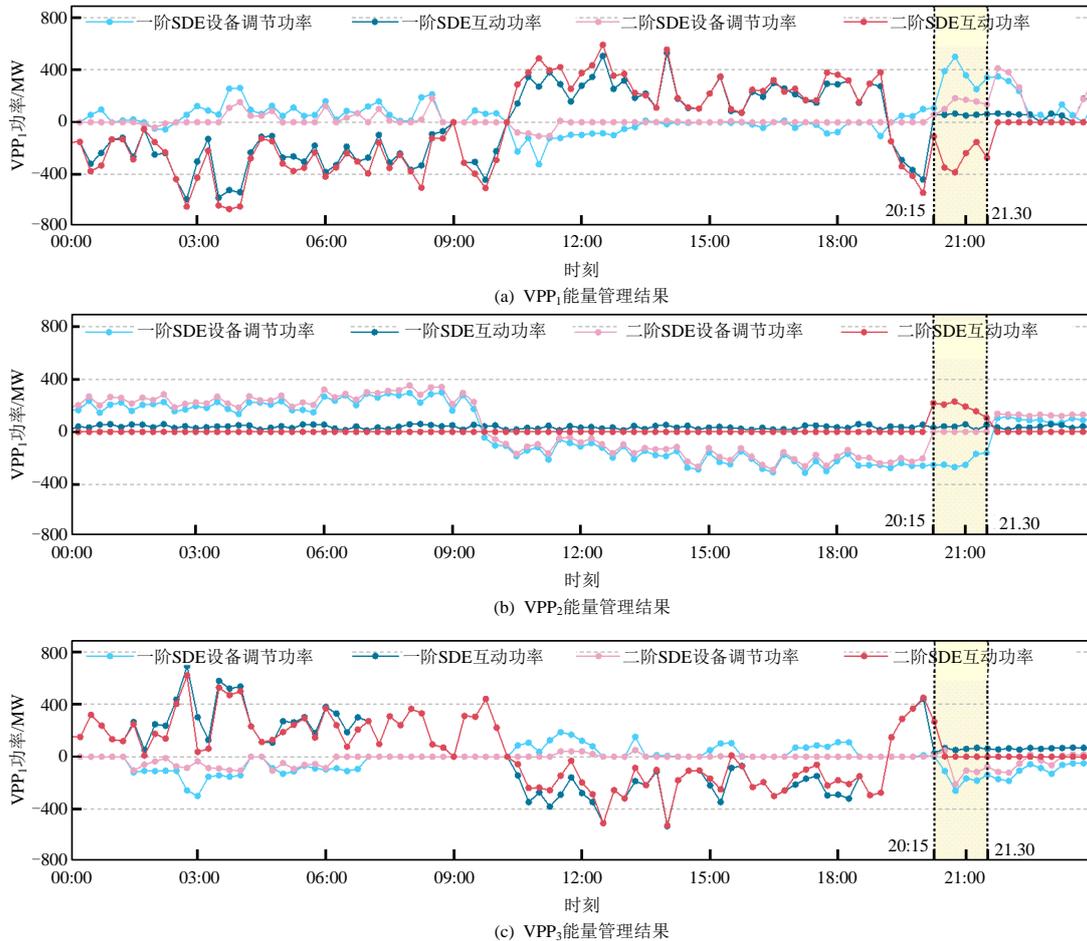


图 6 VPP₃ 策略变化下的 MVPP 系统能量管理结果

Fig. 6 Energy management results of MVPP system under VPP₃ strategy variations

5.3.3 趋优能力对比

一阶和二阶 SDE 驱动下的各 VPP 趋优能力如图 7、8 所示，本文使用每个时刻的调度收益作为评价趋优能力的指标。正常运行时，一阶 SDE 驱动

下 VPP₁—VPP₃ 收益平均值分别为 98.2、170.3、88.2 元；二阶 SDE 驱动时 VPP₁—VPP₃ 收益平均值分别为 131.5、212.5、101.8 元。一阶 SDE 驱动下的设备调节频率更高，所以单调度时刻的收益低于二阶

25.3%、19.6%、17.4%。

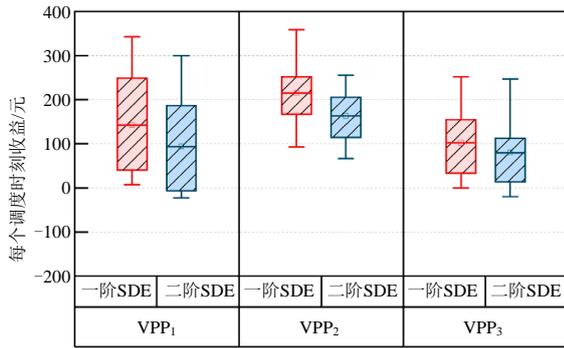


图7 正常运行下MVPP系统各调度时刻的趋优结果
Fig. 7 Optimal results of MVPP system at each scheduling time under normal operation

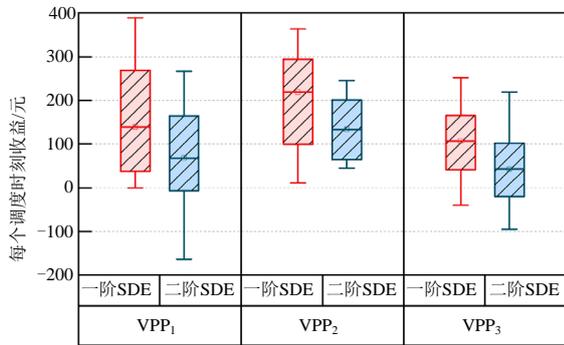


图8 VPP3策略变化下MVPP各调度时刻的趋优结果
Fig. 8 Optimal results of MVPP system at each scheduling time under VPP3 strategy variations

VPP₃策略发生变化,图8记录了各VPP在不同SDE下的趋优结果。一阶SDE驱动下VPP₁—VPP₃收益平均值分别为87.4、165.3、82.2元;二阶SDE驱动时VPP₁—VPP₃收益平均值分别为126.5、218.4、107.7元。VPP₃作为动态决策转变的主体,相较于正常运行时收益平均值提高5.5%,这也是VPP₃选择改变身份的原因。然而可以看到VPP₁收益降低,但由于二阶SDE驱动下与VPP₂的自组织合作联盟使此时的收益下降较少。对于VPP₂而言一阶SDE下的收益平均值降低了2.9%,二阶SDE下反而提升了2.7%。精准的收益状态感知促使VPP₂自趋优性地以合作者的身份应对其他主体策略变化的到来,从而获取了收益提升。

5.4 算法性能对比

为验证本文所提模型与求解算法相对其他方法的优越性,将分布鲁棒优化+CPLEX求解器、无模型+ISAC算法、一阶SDE+ISAC算法、二阶SDE+ISAC算法在VPP₃策略发生改变后的趋优决策效果和决策速度两方面进行对比,对比结果如表1所示。

表1 不同模型与求解算法的趋优效果和速度对比

Table 1 Comparison of effect and speed of different models and solution algorithms

模型	求解算法	趋优收益/元		互动率/%	设备调节率/%	单时刻决策时间/s	
		平均值	标准差			平均值	标准差
分布鲁棒优化	CPLEX	91.47	62.18	14.36	85.61	199.31	14.64
无	ISAC	83.42	79.54	12.07	87.93	4.57	0.84
一阶SDE	ISAC	116.63	123.02	21.07	78.13	2.48	0.34
二阶SDE	ISAC	150.86	63.43	42.37	57.63	2.52	0.27

基于模型的分布鲁棒优化方法依赖于精确的环境模型,且CPLEX每次求解需要重复迭代,因此决策时间最长;无模型的ISAC算法缺乏对策略突然变化数据的训练,因此求解得出的趋优过程中单时刻收益最低,设备调节率最高。本文提出的方法根据随机动力学修正收益结构,通过对求解算法的剪枝实现参数合并,决策时能够快速收敛至最优策略。二阶SDE修正后ISAC求解结果单时刻平均收益提升64.1%,设备调节率下降32.68%。相比无模型的ISAC算法,一阶SDE修正后的模型决策速度提升45.73%,二阶SDE修正后的模型决策速度提升44.9%。可见二阶SDE赋予了VPP较强的自治性能,融合网络的ISAC算法能够大幅提高决策速度。

6 结论

本文提出一种基于二阶随机动力学的多虚拟电厂自趋优能量管理策略,主要结论如下:

1) 聚焦资源可调空间构建VPP资源集群聚合运行模型可以作为解决规模化DERs异质运行特性对能量管理策略影响的一个新思路。

2) 二阶随机动力学方程的引入可以探索其他主体策略突然变化时,VPP自身收益的自发演化信息,从而修正综合收益。形成自趋优机制的模型拥有较好的自治性能,有助于实现多虚拟电厂的高效协同能量管理,提升经济性能。

3) 构建了融合网络的ISAC算法,能够在不失稳定性的情况下拥有更快的收敛速度。

然而,本文中求解算法并未考虑频繁调节对配电网产生的安全约束问题。后续研究将结合安全约束探索多智能体算法收敛条件。

参考文献

[1] 陈启鑫,高洪超,冯成,等.虚拟电厂动态构建与可信量化:理论分析与关键技术[J].电力系统自动化,2022,

- 46(18): 26-36.
- CHEN Qixin, GAO Hongchao, FENG Cheng, et al. Dynamic construction and trustworthy quantification of virtual power plant: Theoretical analysis and key technologies[J]. Automation of Electric Power Systems, 2022, 46(18): 26-36(in Chinese).
- [2] 严兴煜, 高赐威, 陈涛, 等. 数字孪生虚拟电厂系统框架设计及其实践展望[J]. 中国电机工程学报, 2023, 43(2): 604-618.
- YAN Xingyu, GAO Ciwei, CHEN Tao, et al. Framework design and application prospect for digital twin virtual power plant system[J]. Proceedings of the CSEE, 2023, 43(2): 604-618(in Chinese).
- [3] 谢敏, 黄莹, 李弋升, 等. 分布式能源动态聚合于虚拟电厂的演化博弈决策方法和机理分析[J]. 电网技术, 2023, 47(2): 4958-4970.
- XIE Min, HUANG Ying, LI Yisheng, et al. Evolutionary game decision and mechanism analysis of dynamical aggregation of distributed energy resources into virtual power plant[J]. Power System Technology, 2023, 47(2): 4958-4970(in Chinese).
- [4] 薛景润, 施啸寒, 王超, 等. 兼顾物理状态和用户行为的虚拟电厂紧急功率调节能力量化评估[J]. 中国电机工程学报, 2023, 43(8): 2906-2920.
- XUE Jingrun, SHI Xiaohan, WANG Chao, et al. Online evaluation of emergency power regulation capability for virtual power plants considering physical characteristics and user behavior constraints[J]. Proceedings of the CSEE, 2023, 43(8): 2906-2920(in Chinese).
- [5] 李晓舟, 秦文萍, 景祥, 等. 计及不确定风险和多主体协同的虚拟电厂参与主辅市场联合优化策略[J/OL]. 电网技术: 1-16[2023-10-27]. <https://doi.org/10.13335/j.1000-3673.pst.2023.1309>.
- LI Xiaozhou, QIN Wenping, JING Xiang, et al. Joint optimization strategy for virtual power plant participating in main and auxiliary markets considering uncertain risks and multi-agent coordination[J]. Power System Technology: 1-16[2023-10-27]. <https://doi.org/10.13335/j.1000-3673.pst.2023.1309>(in Chinese).
- [6] 范帅, 何光宇, 郑湘明, 等. 基于在线分布式优化的虚拟电厂自趋优运行方法研究[J]. 中国电机工程学报, 2023, 43(13): 4935-4949.
- FAN Shuai, HE Guangyu, ZHENG Xiangming, et al. Research on online distributed optimization-based self-approaching optimization operation method of virtual power plant[J]. Proceedings of the CSEE, 2023, 43(13): 4935-4949(in Chinese).
- [7] 王文悦, 刘海涛, 季宇. 虚拟电厂可调空间统一建模及其参与调峰市场的优化运行策略[J]. 电力系统自动化, 2022, 46(18): 74-82.
- WANG Wenyue, LIU Haitao, JI Yu. Unified modeling for adjustable space of virtual power plant and its optimal operation strategy for participating in peak-shaving market [J]. Automation of Electric Power Systems, 2022, 46(18): 74-82(in Chinese).
- [8] 侯慧, 甘铭, 吴细秀, 等. 考虑移动氢能存储的港口多能微网两阶段分布鲁棒优化调度[J]. 中国电机工程学报, 2024, 44(8): 3078-3092.
- HOU Hui, GAN Ming, WU Xixiu, et al. Two-stage distributionally robust optimal scheduling for port multi-energy microgrid considering mobile hydrogen energy storage[J]. Proceedings of the CSEE, 2024, 44(8): 3078-3092(in Chinese).
- [9] 张亚超, 郑峰, 舒胜文, 等. 考虑多重不确定性的电气-交通网络耦合系统数据驱动鲁棒优化调度[J]. 中国电机工程学报, 2021, 41(13): 4450-4461.
- ZHANG Yachao, ZHENG Feng, SHU Shengwen, et al. A data-driven robust optimization scheduling of coupled electricity-gas-transportation systems considering multiple uncertainties[J]. Proceedings of the CSEE, 2021, 41(13): 4450-4461(in Chinese).
- [10] ZHAO Zhuoli, GUO Juntao, LUO Xi, et al. Distributed robust model predictive control-based energy management strategy for islanded multi-microgrids considering uncertainty[J]. IEEE Transactions on Smart Grid, 2022, 13(3): 2107-2120.
- [11] GLEESON J P. Binary-state dynamics on complex networks: Pair approximation and beyond[J]. Physical Review X, 2013, 3(2): 021004.
- [12] ZHUKOV D, KHVATOVA T, MILLAR C, et al. Modelling the stochastic dynamics of transitions between states in social systems incorporating self-organization and memory[J]. Technological Forecasting and Social Change, 2020, 158: 120134.
- [13] ZHUKOV D, KHVATOVA T, MILLAR C, et al. Beyond big data - new techniques for forecasting elections using stochastic models with self-organization and memory[J]. Technological Forecasting and Social Change, 2022, 175: 121425.
- [14] 陈中, 陈嘉琛, 万玲玲. 基于随机演化动力学的多微网-配电网自组织协同调节策略[J]. 电力系统自动化, 2023, 47(2): 24-33.
- CHEN Zhong, CHEN Jiachen, WAN Lingling. Self-organized cooperative regulation strategy for multi-microgrid-distribution network based on stochastic evolutionary dynamics[J]. Automation of Electric Power Systems, 2023, 47(2): 24-33(in Chinese).
- [15] WOOD A J, WOLLENBERG B F. Power generation, operation, and control[M]. 2nd ed. Hoboken, NJ, USA: Wiley, 2012.

- [16] ANTONIADOU-PLYTARIA K, STEEN D, TUAN L A, et al. Scenario-based stochastic optimization for energy and flexibility dispatch of a microgrid[J]. IEEE Transactions on Smart Grid, 2022, 13(5): 3328-3341.
- [17] CHEN Bo, WANG Jianhui, LU Xiaonan, et al. Networked microgrids for grid resilience, robustness, and efficiency: a review[J]. IEEE Transactions on Smart Grid, 2021, 12(1): 18-32.
- [18] YI Zhongkai, XU Yinliang, WANG Xue, et al. An improved two-stage deep reinforcement learning approach for regulation service disaggregation in a virtual power plant[J]. IEEE Transactions on Smart Grid, 2022, 13(4): 2844-2858.
- [19] 张超, 赵冬梅, 季宇, 等. 基于改进深度 Q 网络的虚拟电厂实时优化调度[J]. 中国电力, 2024, 57(1): 91-100. ZHANG Chao, ZHAO Dongmei, JI Yu, et al. Real time optimal dispatch of virtual power plant based on improved deep Q network[J]. Electric Power, 2024, 57(1): 91-100(in Chinese).
- [20] 张继行, 张一, 王旭, 等. 基于多代理强化学习的多新型市场主体虚拟电厂博弈竞价及效益分配策略[J]. 电网技术, 2024, 48(5): 1980-1991. ZHANG Jixing, ZHANG Yi, WANG Xu, et al. Game bidding and benefit allocation strategies for virtual power plants with multiple new market entities based on multi-agent reinforcement learning[J]. Power System Technology, 2024, 48(5): 1980-1991(in Chinese).
- [21] FENG Bin, LIU Zhuping, HUANG Gang, et al. Robust federated deep reinforcement learning for optimal control in multiple virtual power plants with electric vehicles[J]. Applied Energy, 2023, 349: 121615.
- [22] SUBRAMANIAN L, DEBUSSCHERE V, GOOI H B. A distributed model predictive control framework for grid-friendly distributed energy resources[J]. IEEE Transactions on Sustainable Energy, 2021, 12(1): 727-738.
- [23] ZHUKOVSKII M E, RAIGORODSKII A M. Random graphs: models and asymptotic characteristics[J]. Russian Mathematical Surveys, 2015, 70(1): 33-81.
- [24] HE Jialu, WANG Jianwei, YU Fengyuan, et al. The persistence and transition of multiple public goods games resolves the social dilemma[J]. Applied Mathematics and Computation, 2022, 418: 126858.
- [25] ZHENG Jiaohao, KURT M N, WANG Xiaodong. Stochastic integrated actor-critic for deep reinforcement learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 35(5): 1-13.
- [26] ZHANG Zhenyuan, CHEN Zihan, LEE W J. Soft actor-critic algorithm featured residential demand response strategic bidding for load aggregators[J]. IEEE Transactions on Industry Applications, 2022, 58(4): 4298-4308.

附录

<http://ntps.epri.sgcc.com.cn/djgxcb/CN/10.13334/i.0258-8013.pcsee.232507>



陈嘉琛

在线出版日期: 2024-05-15。

收稿日期: 2023-11-14。

作者简介:

陈嘉琛(2000), 男, 博士研究生, 主要从事分布式资源聚合运行、多虚拟电厂优化等研究工作, 230228741@seu.edu.cn;

*通信作者: 陈中(1975), 男, 研究员, 博士生导师, 主要从事人工智能在电力系统中的应用、电动汽车电网智能互动, zhongchen@seu.edu.cn。

(编辑 乔宝榆, 李新洁)

Energy Management Strategy for Multi-virtual Power Plants with Self-optimization Based on Second-order Stochastic Dynamics

CHEN Jiachen, CHEN Zhong*, LI Bingrong, LIU Wenyu, PAN Jundi¹

(School of Electrical Engineering, Southeast University)

KEY WORDS: multi-virtual power plant; self-optimization; aggregate operation model; second-order stochastic dynamics; multi-agent deep reinforcement learning

The presence of numerous stochastic elements in distributed energy resources (DERs) leads to frequent changes in Multi-Virtual Power Plant (MVPP) when it comes to the strategy of individual VPPs. For one given entity, understanding the trend of the impact on its own returns when perceiving sudden changes in the strategies of other entities and rapidly adjusting its own optimization strategy is a critical issue that urgently needs to be addressed.

This paper proposes a self-trending optimization strategy for MVPPs based on second-order stochastic dynamics, aiming to enhance the autonomy of VPPs in responding to changes in the strategies of other entities. Firstly, addressing the heterogeneous operational characteristics of DERs, this paper focuses on the adjustable space of resources to construct a clustered operational model for VPP resources. Secondly, the stochastic nature of VPP strategy transitions is depicted based on the theory of random graphs. Thirdly, second-order stochastic dynamic equations are used to explore its spontaneous evolutionary information to adjust the comprehensive profit of VPPs when other entities' strategies changed. Moreover, the adjusted profit is used as the true reward function for the Integrated Soft Actor-Critic (ISAC) deep reinforcement learning decision model to establish multi-agent distributed solution framework. Finally, multiple algorithm

comparison experiments are designed to validate the self-trending performance of the proposed strategy in this paper.

The SAC algorithm incorporates maximum entropy principles when facing stochastic changes, aiding the agent in selecting actions from the action space as randomly as possible. The gradient of the entropy regularization coefficient α is

$$J(\alpha) = \mathbb{E}_{a_k \sim \pi_k} [-\alpha \log \pi_k(a_k | s_k) - \alpha \bar{H}] \quad (1)$$

where a_k is the action of the agent, s_k is the state of the system, π_k is the action strategy and \bar{H} is the entropy term. The framework is shown in Fig. 1.

Compared to other methods, the second-order stochastic dynamic equations introduced in this paper can explore the spontaneous evolution information of VPP's own benefits when other agent strategies suddenly change, thereby correcting the comprehensive profit. The model with self-trending optimization mechanism has better autonomy performance, which contributes to efficient collaborative energy management of MVPP and improves economic performance. Additionally, the ISAC algorithm constructed in this paper merges two Q-networks and shares inputs with the policy network and the training network, which can achieve faster convergence speed without sacrificing stability.

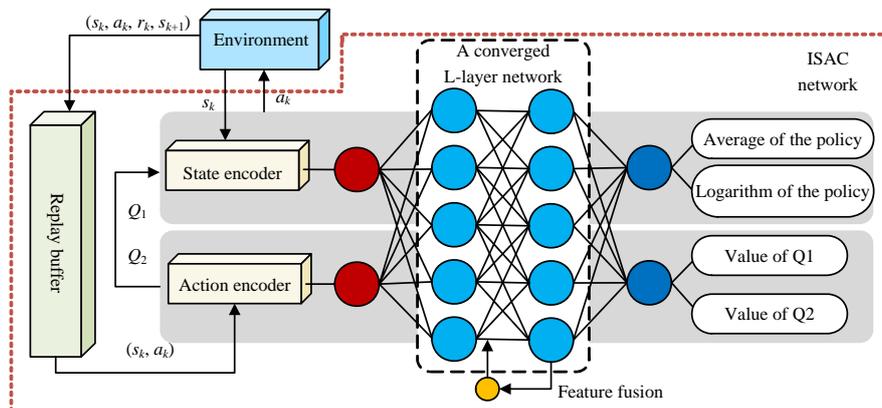


Fig. 1 ISAC algorithm decision-making framework