DOI: 10.19783/j.cnki.pspc.250435

面向频率稳定的基于预生成经验池驱动的紧急切负荷智能在线决策

李成翔[1,2,3]，龚楷程[4]，朱益华[1,2,3]，梁卓航[1,2,3]，兰宇田[4]，韦善阳[4]，姚 伟[4]

(1. 直流输电技术全国重点实验室(南方电网科学研究院有限责任公司)，广东 广州 510663；2. 国家能源大电网技术研发(实验)中心，广东 广州 510663；3. 广东省新能源电力系统智能运行与控制重点实验室，广东 广州 510663；4. 强电磁技术全国重点实验室(华中科技大学电气与电子工程学院)，湖北 武汉 430074)

摘要：随着电网规模的扩展和新能源渗透率的提高，电力系统频率稳定问题已成为当前研究的重点。紧急切负荷是应对电力系统频率失稳的有效控制手段，但传统紧急切负荷控制采用离线整定-在线匹配模式已难以适应复杂多变的故障场景。提出一种基于深度强化学习的紧急切负荷在线决策方法。首先，提出基于马尔可夫决策过程(Markov decision process, MDP)的紧急切负荷决策建模方法，同时采用分支竞争 Q 网络(branch dueling Q-network, BDQ)应对高维切负荷决策空间。其次，针对传统强化学习训练中时间与计算成本高昂的问题，通过解耦样本采集与模型训练环节，采用预生成经验池驱动的集中训练策略，实现智能体的高效训练。最后，基于 10 机 39 节点系统的算例验证表明，所提算法在决策有效性上较传统强化学习方法提升 4.94%，训练所需时间仅为传统方法的 8.82%。

关键词：紧急切负荷；深度强化学习；频率安全；在线决策

Frequency stability-oriented intelligent online decision-making for emergency load shedding based on a pre-generated experience pool

LI Chengxiang[1,2,3], GONG Kaicheng[4], ZHU Yihua[1,2,3], LIANG Zhuohang[1,2,3], LAN Yutian[4], WEI Shanyang[4], YAO Wei[4]

(1. State Key Laboratory of HVDC, Electric Power Research Institute, China Southern Power Grid, Guangzhou 510663, China; 2. National Energy Power Grid Technology R&D Centre, Guangzhou 510663, China; 3. Guangdong Provincial Key Laboratory of Intelligent Operation and Control for New Energy Power System, Guangzhou 510663, China; 4. State Key Laboratory of Advanced Electromagnetic Engineering and Technology (School of Electrical and Electronic Engineering, Huazhong University of Science and Technology), Wuhan 430074, China)

Abstract: With the expansion of power grids and the increasing penetration of renewable energy, frequency stability in power systems has become a key research focus. Emergency load shedding is an effective control strategy to mitigate frequency instability; however, the traditional offline-setting and online-matching approach is increasingly inadequate for complex and evolving fault scenarios. To address this issue, a deep reinforcement learning-based method is proposed to achieve online decision-making for emergency load shedding. First, a decision modeling approach based on Markov decision process (MDP) is developed, and a branch dueling Q-network (BDQ) is introduced to handle the high-dimensional load shedding decision space. Furthermore, to overcome the high computational and time costs in traditional reinforcement learning training, a centralized training strategy driven by a pre-generated experience pool is adopted by decoupling sample collection from model training, thereby enabling efficient agent training. Finally, simulation results on a 10-machine 39-bus system demonstrate that the proposed algorithm improves decision effectiveness by 4.94% compared to traditional reinforcement learning methods, while reducing training time to only 8.82% of that of conventional approaches.

This work is supported by the National Natural Science Foundation of China (No. U22B20111).

Key words: emergency load shedding; deep reinforcement learning; frequency security; online decision-making

基金项目：国家自然科学基金项目资助(U22B20111)；南方电网科学研究院科技项目资助(SEPRI-K233009)

0　引言

在“双碳”战略背景下，构建以新能源为主体的新型电力系统成为实现低碳转型的关键路径。大规模新能源并网将成为我国未来电力系统的重要特征[1-2]，这对电力系统的频率稳定性提出了多方面的挑战[3]。新能源发电的随机性易引发频率扰动[4]，而电力电子设备缺乏惯量支撑能力[5]，进一步放大频率失稳风险。电网可能因直流闭锁[6]、换流器故障或新能源集群脱网[7]等事故出现频率失稳甚至崩溃。紧急切负荷作为安全稳定控制的第二道防线，通过快速切除负荷补偿系统功率缺额，是维持频率稳定的重要紧急控制手段。

当前电力系统紧急切负荷决策主要采用离线整定-在线匹配的控制框架[8-9]，目前研究大都聚焦于切负荷决策的快速自动化生成，包括切负荷的地点、时间以及切负荷量[10]。传统方法将问题建模为最优控制模型，结合灵敏度分析辅助决策。文献[11]以切负荷代价最小为优化目标，同时利用轨迹灵敏度将安全约束线性化转化为线性规划求解。文献[12]通过引入切负荷对暂态稳定裕度的灵敏度，将非线性优化问题近似为线性规划问题。上述方法计算量大、适用工况单一，难以应对复杂多变的故障场景。

近年来随着人工智能技术的进步，深度学习特别是深度强化学习(deep reinforcement learning, DRL)为紧急切负荷决策提供了新思路。文献[13]基于改进的 AlexNet 深度卷积神经网络，预测切负荷动作的灵敏度以辅助控制措施的选择。文献[14]将深度强化学习应用于电压稳定紧急切机控制。文献[15]基于深度强化学习，实现了电压稳定的紧急切负荷控制。上述方法主要聚焦于紧急决策的离线制定，而未充分探究强化学习方法的在线决策潜力。随着电网规模增长以及系统新能源渗透率的提高，离线整定所需枚举的故障工况成倍增长[16]。在线匹配的应用模式依赖于检索相似工况，对新型复杂工况适应性不足。基于实时运行状态的切负荷在线决策，将成为实现精确、自适应紧急控制的关键突破方向。

深度强化学习在电力系统紧急控制中具有巨大的应用潜力[17-18]，但是在智能体的训练上存在瓶颈。现有训练范式大都采用电力系统暂态仿真作为交互环境[19]。强化学习基于与环境交互进行学习的特点，导致其训练过程中需要频繁调用电力系统仿真。特别是在超参数优化阶段，仿真计算与模型迭代的深度耦合导致训练过程产生海量计算需求，造成时间成本激增与硬件资源超载，严重制约了智能体训练的效率。如何优化交互过程，提升训练效率，已成为强化学习领域亟需解决的核心问题之一。针对上述问题，文献[20]采用融合了去除重复行动和消极行动的知识提高了样本的探索效率。文献[21]提出并行增强随机搜索算法(parallel augment random search, PARS)在大规模强化学习训练过程中执行结构化的、更有效的参数空间探索。上述方法一定程度上提升了训练的效率，然而并没有从根本上解决强化学习反复调用仿真的问题。

针对上述挑战，本文提出基于预生成经验池驱动的紧急切负荷智能在线决策方法。首先构建了基于强化学习算法的紧急切负荷决策框架，其中包括智能体决策流程、切负荷决策特征的选择、以及基于电力系统暂态仿真的智能体训练环境。其次，针对紧急切负荷控制指令的离散性，设计了离散动作空间，同时引入分支竞争 Q 网络(branch dueling Q-network, BDQ)算法以应对动作空间维度高的问题。最后，针对强化学习训练需要频繁调用电力系统仿真以及训练成本高昂、调参困难的问题，本文针对经验回放机制进行改进，提出基于预生成经验池驱动的集中训练策略，有效减少了训练的时间与计算成本。基于 10 机 39 节点系统的实验表明，所提方法在决策效果更优的同时，训练耗时仅为传统强化学习方法的 8.82%，综合性能显著提升。

1　紧急切负荷智能体决策训练框架

本文所提出的基于强化学习方法的紧急切负荷智能体决策与训练框架如图 1 所示，主要分为智能体集中训练与在线决策两部分。

1) 智能体集中训练

本文改进传统强化学习经验回放机制，采用基于预生成经验池驱动的集中训练策略。传统智能体训练交替执行两个步骤：智能体通过环境交互生成经验样本存入回放单元；从回放单元采样进行模型训练。

对于训练过程改进的核心在于：通过预计算生成仿真经验池替代传统实时交互采样，实现样本生成与模型训练的解耦。具体实施分为两个阶段：第一阶段，由强化学习智能体对预想故障样本进行预决策，并输入电力系统暂态仿真，从而获得暂态稳定仿真数据，生成经验样本存入经验池；第二阶段，基于预生成经验池训练强化学习智能体，并通过仿真验证决策效果以调整超参数。

2) 在线决策

将训练完成的深度强化学习模型部署至在线系统。通过 SCADA 系统实时监测电网状态，当检测到严重故障并触发切负荷需求时，将实时状态数据

输入训练完成的智能体，由其生成控制指令。紧急切负荷装置依据输出的最优动作方案执行切负荷操作，从而实现频率稳定控制。

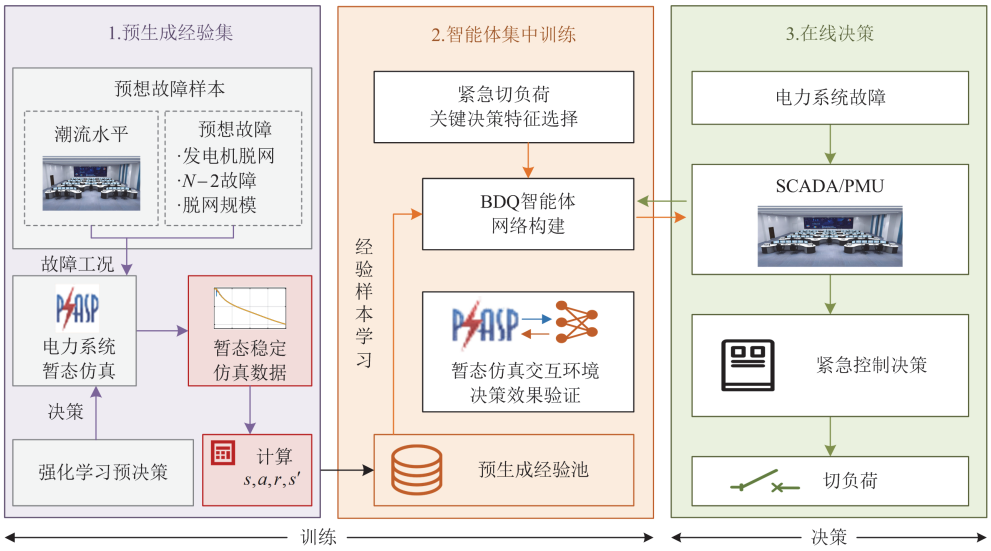


图 1 智能体紧急切负荷决策与训练框架

Fig. 1 A framework for emergency load cutting decision making and training of agent

2 基于深度强化学习的紧急切负荷决策问题建模

2.1 适用于紧急切负荷决策的强化学习 MDP 过程设计

强化学习是机器学习的一个重要分支，其训练智能体通过与环境持续交互来学习以完成特定目标[22]。深度强化学习结合了强化学习与深度学习，借助神经网络处理高维复杂的状态空间，从而使得强化学习(reinforcement learning, RL)算法能够适用于更复杂的问题[23-24]。

将深度强化学习应用于紧急切负荷决策，首先需要将切负荷决策问题抽象成马尔可夫决策过程(Markov decision process, MDP)[25]。一个 MDP 过程通常由状态空间、动作空间、奖励函数、状态转移等构成。下面介绍适用于紧急切负荷的 MDP 构建[26]。

针对电力系统紧急切负荷场景的特殊性，本研究对传统多步马尔可夫决策过程框架进行了改进，提出了一种单步决策机制。该机制通过限定每回合仅允许执行单次切负荷动作，从而避免了多步决策可能引发的延迟响应风险，确保在故障发生时能立即实施足量的控制。改进后的单步 MDP 可视为常规多步决策过程的一种简化特例，更贴合电力系统频率崩溃场景对时效性的要求。

2.1.1 状态、动作空间设计

强化学习状态 $\boldsymbol{s}$ 是对当前时刻环境的概括，智能体切负荷的决策便是基于策略根据当前环境生成的。因此合理设计智能体状态空间对于决策效果至关重要。环境特征 $\boldsymbol{f}$ 以及状态 $\boldsymbol{s}$ 是强化学习的两个重要概念。其中，环境特征 $\boldsymbol{f}$ 是对环境的全面描述，智能体观测环境特征，从中提取出用于决策的特征子集，即构成了强化学习的状态 $\boldsymbol{s}$。从环境特征 $\boldsymbol{f}$ 中提取状态 $\boldsymbol{s}$ 的过程可以用一个映射函数表示。

$$\boldsymbol{s}=\phi(\boldsymbol{f}) \tag{1}$$

本文通过紧急切负荷控制来维持系统的频率稳定，因此首先需要筛选出能反映系统频率稳定的系统特征 $\boldsymbol{f}$。

考虑到电力系统的频率稳定与系统有功功率平衡强相关，因此选择各发电机节点频率 $\boldsymbol{f}_{\mathrm{G}}$、各发电机节点的有功功率 $\boldsymbol{P}_{\mathrm{G}}$ 与负荷节点的有功功率 $\boldsymbol{P}_{\mathrm{L}}$。考虑到故障后发电机的暂态过程，发电机节点的有功出力可能并不能真实反映电力系统的有功功率平衡情况，因此增加发电机节点的机械功率 $\boldsymbol{P}_{\mathrm{m}}$，以此更好地描述系统的频率稳定情况。筛选得到电力系统特征可以表示为

$$\boldsymbol{f}=(\boldsymbol{f}_{\mathrm{G}},\boldsymbol{P}_{\mathrm{G}},\boldsymbol{P}_{\mathrm{m}},\boldsymbol{P}_{\mathrm{L}})=\begin{bmatrix} f_{\mathrm{G1}},\cdots,f_{\mathrm{G}n_{\mathrm{G}}} \\ P_{\mathrm{G1}},\cdots,P_{\mathrm{G}n_{\mathrm{G}}} \\ P_{\mathrm{G1}},\cdots,P_{\mathrm{G}n_{\mathrm{G}}} \\ P_{\mathrm{L1}},\cdots,P_{\mathrm{L}n_{\mathrm{L}}} \end{bmatrix} \tag{2}$$

式中：n_{G} 为系统发电机数；n_{L} 为系统负荷数。系统特征包含 $[0,T]$ 仿真时间内所有时序数据。

在选定能够反映系统频率的环境特征后，需要考虑环境特征 $\boldsymbol{f}$ 与状态 $\boldsymbol{s}$ 的映射 ϕ，即如何从环境特征 $\boldsymbol{f}$ 中提取出状态 $\boldsymbol{s}$，这与实际控制需求有关。如前所述，本文所提的基于深度强化学习的紧急切负荷决策采用离线训练-在线决策的控制框架。虽然在离线训练过程中，可以通过电力系统仿真得到故障后全时域的仿真数据，但是考虑到在线应用的情景，电力系统发生故障需要智能体进行紧急切负荷决策时，智能体所能获取的电力系统特征 $\boldsymbol{f}$ 仅为控制投入前的信息，智能体输入的状态 $\boldsymbol{s}$ 也只能由该控制投入前的电力系统特征 $\boldsymbol{f}$ 构成。因此选择 t 时刻控制投入时的电力系统特征 $\boldsymbol{f}$ 构成强化学习状态 $\boldsymbol{s}$。

$$\boldsymbol{s}=(\boldsymbol{f}_{\mathrm{G},t},\boldsymbol{P}_{\mathrm{G},t},\boldsymbol{P}_{\mathrm{m},t},\boldsymbol{P}_{\mathrm{L},t})=\begin{bmatrix} f_{\mathrm{G1},t},\cdots,f_{\mathrm{G}n_{\mathrm{G}},t} \\ P_{\mathrm{G1},t},\cdots,P_{\mathrm{G}n_{\mathrm{G}},t} \\ P_{\mathrm{m1},t},\cdots,P_{\mathrm{m}n_{\mathrm{G}},t} \\ P_{\mathrm{L1},t},\cdots,P_{\mathrm{L}n_{\mathrm{L}},t} \end{bmatrix} \tag{3}$$

动作 $\boldsymbol{a}$ 是智能体依据当前状态所做出的决策。动作空间是所有可能动作的集合。本文将强化学习用于紧急切负荷决策，因此各智能体的动作即是决策的各机组切负荷量，表示为

$$\boldsymbol{a}=(L_s^1,L_s^2,\cdots,L_s^i) \tag{4}$$

式中：L_s^i 表示第 i 个负荷的切除量。

2.1.2 基于暂态仿真的状态转移过程

在本文中，采用电力系统暂态仿真作为强化学习的环境，相应的系统状态转移由电力系统仿真实现，在智能体做出决策后，将各负荷节点的切负荷量输入电力系统仿真软件，获取上述电力系统特征，即发电机节点频率 $\boldsymbol{f}_{\mathrm{G}}$、发电机节点的有功功率 $\boldsymbol{P}_{\mathrm{G}}$、发电机节点的机械功率 $\boldsymbol{P}_{\mathrm{m}}$ 以及负荷节点的有功功率 $\boldsymbol{P}_{\mathrm{L}}$ 的全时域仿真结果，截取对应时刻的电力系统特征作为电力系统的下一个状态返回给智能体，从而实现状态转移。因此需要构建强化学习智能体与电力系统暂态仿真的交互框架，使得智能体能够合理高效地与环境进行交互。图 2 给出了基于暂态稳定仿真的智能体与环境交互框架。

在初始化切负荷决策 u_0 后，选择某一训练样本，依据其预想故障 F 得到仿真条件 (F,u_0) 进行暂态仿真得到仿真数据 f_0。仿真数据 f_0 是一串时域数据，在事件驱动紧急切负荷的具体应用情景下，选择故障发生后 0.1 s 的数据作为状态 $\boldsymbol{s}$ 输入，智能体在接收到状态 $\boldsymbol{s}$ 输入后做出决策 $\boldsymbol{u}$。依据新的仿真条件 $(F,\boldsymbol{u})$ 再次进行暂态仿真得到仿真数据 f_1。选择仿真数据 f_1 中控制投入后 2 s 的数据作为状态 $\boldsymbol{s}'$ 返回给智能体，同时取仿真数据 f_1 末尾数据得到系统恢复频率，进而计算奖励 r。

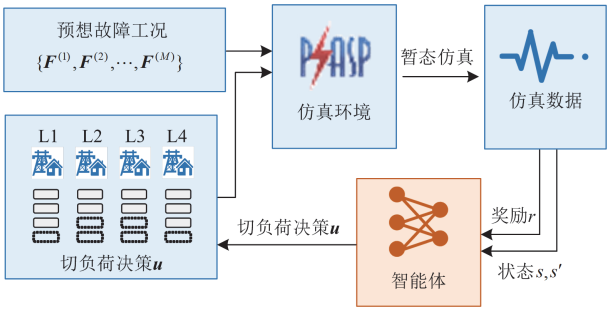


图 2 智能体与环境交互框架

Fig. 2 Agent-environment interaction framework

2.1.3 奖励函数设计

奖励 r 是系统反馈给智能体的对于智能体动作效果的评价，其设计需与决策目标深度耦合[27]。紧急切负荷控制的首要目标是保障系统稳定，因此奖励函数的设计首先考虑决策的有效性。

1) 决策成功奖励：当智能体执行切负荷操作后，系统频率能够保持稳定(即未发生失稳)，则给予智能体一个固定奖励。该奖励旨在引导智能体优先学习保障系统稳定性的决策，提升切负荷操作的成功率。

在确保系统稳定的基础上，奖励函数进一步关注决策的精确性，具体通过以下两个指标量化：

2) 切负荷量惩罚：根据当前决策实际切除的负荷大小给予惩罚，切除负荷越多，惩罚值越大，旨在鼓励智能体在保证稳定的前提下尽可能减少切除的负荷量。

3) 频率收敛奖励：根据系统最终恢复的惯性中心频率与最低允许恢复频率(49.5 Hz)之间的绝对误差，乘以一定系数后给予奖励，用以鼓励智能体提升频率恢复的准确性。

综上，奖励函数 r 可由决策成功奖励、切负荷量惩罚和频率收敛奖励 3 部分组成，如式(5)所示。

$$r=r_{\mathrm{success}}-\lambda_1\sum_{i=1}^{n_{\mathrm{L}}}L_s^i+\lambda_2\left|f_{\mathrm{COI\cdot stable}}-49.5\right| \tag{5}$$

式中：r_{success} 为决策成功奖励；$f_{\mathrm{COI\cdot stable}}$ 为系统恢复频率(以惯性中心频率计)；λ_1 为切负荷的惩罚系数；λ_2 为频率收敛的奖励系数。奖励系数的设定需通过调参确定，基本思路是首先赋予较大的决策成功奖励以保证决策成功率，随后逐步增加切负荷量惩罚和频率收敛奖励的权重，以兼顾决策的精确性。

2. 2 BDQ 强化学习智能体设计

2.2.1 BDQ 智能体神经网络设计

在应对切负荷控制等离散动作空间问题时，深

度 Q 网络(deep Q-network, DQN)、双 Q 网络(double Q-learning, DDQN)等算法容易因动作空间的高维导致智能体训练困难。BDQ 网络算法是一种强化学习算法，专门用于处理具有多个竞争性分支的问题。该算法可视为 DQN 的扩展[28-29]。

BDQ 算法旨在解决多竞争性分支并存的问题，其中每个分支代表不同的决策路径或策略选择。其核心是智能体需基于当前环境状态及各分支的竞争关系，选择最优决策以实现目标优化或协同控制。本文将紧急频率控制中的切负荷问题建模为多分支竞争问题，将各可切除负荷视为竞争性分支，通过负荷间的协同控制获取最优切负荷方案。

基于 BDQ 算法的神经网络结构如图 3 所示，主要由共享网络层、分支网络层和状态价值分支 3 部分构成。共享网络层负责提取输入状态的特征，这些特征将供后续分支网络层和状态价值分支共同使用。根据 MDP 建模选取的 4 个状态张量(发电机节点频率、发电机有功功率、发电机机械功率、负荷有功功率)，设计两步特征提取方式：首先对各状态张量进行独立特征提取，随后将输出特征拼接并通过全连接层形成共享网络层。

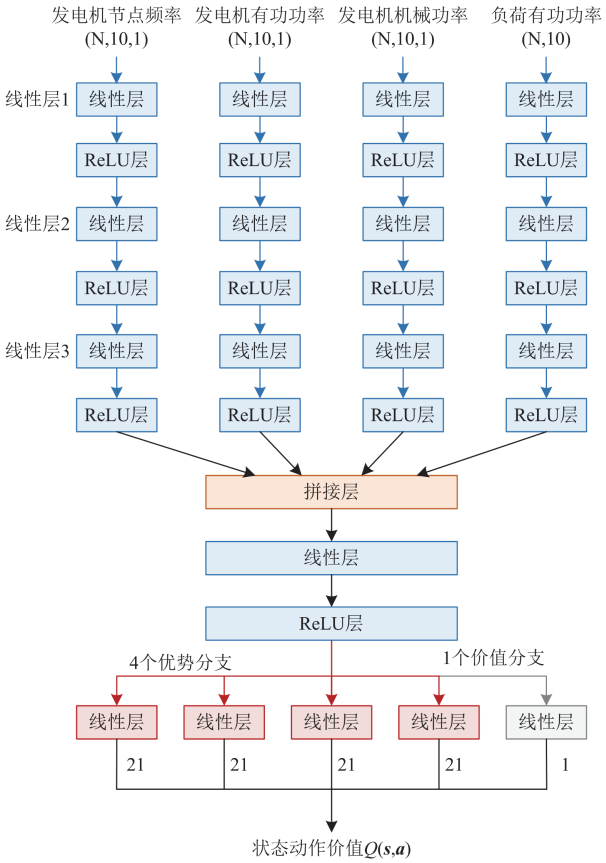


图 3 基于 BDQ 算法的神经网络

Fig. 3 Neural network based on BDQ algorithm

分支网络层中的每个分支对应一个可切负荷节点，与状态价值分支共享特征参数。BDQ 算法的核心机制在于通过分支网络和状态价值分支协同计算 Q 值，其创新点在于将状态价值函数 $V(\boldsymbol{s})$ 与各动作维度的优势函数 $A(a_i)$ 分离估计，然后计算得到最终的 Q 值。

$$Q(\boldsymbol{s},\boldsymbol{a})=V(\boldsymbol{s})+\sum_{i=1}^{n}\left(A_i(\boldsymbol{s},a_i)-\frac{1}{|\mathcal{A}_i|}\sum_{a_{i'}\in\mathcal{A}_i}A_i(\boldsymbol{s},a_{i'})\right) \tag{6}$$

式中：$\boldsymbol{a}$ 是动作向量，每个 a_i 对应一个动作维度；$\mathcal{A}_i$ 是第 i 个动作维度的动作空间；$|\mathcal{A}_i|$ 表示第 i 个动作维度所有可选的动作数量。

2.2.2 BDQ 算法下动作与决策空间的降维

强化学习算法按动作空间是否连续可以分为离散控制与连续控制，鉴于切负荷动作实际上是切除该负荷对应的出线，是一个离散控制问题，因此本文采用离散动作空间进行建模。

在处理复杂场景时，传统基于离散动作空间的强化学习算法面临着维度灾难。以 10 机 39 节点系统为例，系统配置 4 个可切负荷节点，每个节点设置 0~100%负荷量的 21 级调节梯度(步长 5%)，传统算法下需构建 21^4 维动作空间，这意味着相应的神经网络输出层共有 19 481 个输出神经元，显著增加了模型复杂度与训练难度。当系统规模进一步扩大时，该维度还会以指数趋势增长。

BDQ 网络中每个可切负荷都可视为一个独立的动作维度，拥有独立分支网络层，各个分支网络层分别估计该可切负荷各个切负荷量的 Q 值。因此同样的 10 机 39 节点电力系统中，采用 BDQ 方式构建的神经网络输出神经元个数仅为 $4\times21=84$ 个，实现了两个数量级的降维。同时也解决了系统进一步扩大带来的维度爆炸问题。

3　预生成经验池驱动的智能体集中训练

基于第 2 节对紧急切负荷问题的强化学习建模，现已能够实现对智能体的训练。然而，在常规强化学习方法中，智能体需不断与环境交互，在每一步决策后实时收集经验样本(包括状态 $\boldsymbol{s}$、动作 $\boldsymbol{a}$、奖励 r 和下一个状态 $\boldsymbol{s}'$)，动态构建经验集。这意味着每次训练，甚至每次调参时，都需要频繁调用电力系统的暂态仿真以实时采集新样本，导致计算和时间成本大幅增加，严重制约了训练和调参效率。

上述问题的根源在于，经验收集与网络优化两过程高度耦合。为此，本文提出对二者进行解耦：即先通过批量仿真集中生成经验集，再基于该经验集进行高效的集中训练，从而显著减少仿真调用次数，提升整体训练效率。

3.1 异策略强化学习的经验回放机制

与 DQN、DDQN 等算法类似，BDQ 算法同属异策略强化学习框架，因此同样采用经验回放机制进行智能体训练。图 4 为基于经验回放机制的 BDQ 智能体训练流程。

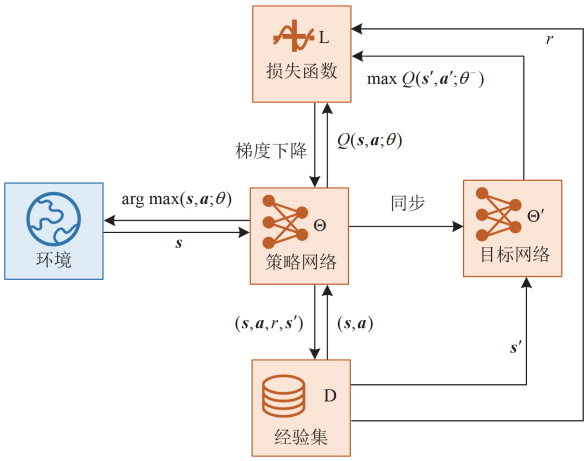


图 4 BDQ 算法流程示意图

Fig. 4 Schematic diagram of BDQ algorithm flow

在决策尝试过程中，假设智能体在状态 $\boldsymbol{s}$ 下做出动作 $\boldsymbol{a}$，环境接受动作 $\boldsymbol{a}$ 后状态转移至 $\boldsymbol{s}'$，同时得到该步的奖励 r。将上述 $(\boldsymbol{s},\boldsymbol{a},r,\boldsymbol{s}')$ 作为经验存入经验集。每回合动作结束后，从经验集中抽取一定经验样本进行时序差分(temporal-difference, TD)计算，利用所得损失对网络进行梯度下降，即完成了一次网络的训练优化。上述过程即是基于经验回放机制的异策略强化学习智能体训练过程。

可以发现在经验回放机制下，异策略深度强化学习的训练过程是利用经验集中存储的经验训练一个神经网络，本质上也可以归结为一个深度学习。从上述流程中也可以看到，经验集的数据依赖于强化学习框架下智能体的决策过程，若能预先生成一个经验集合取代经验集，则可以将上述训练神经网络的过程单独抽离出来，集中训练强化学习智能体。

3.2 强化学习智能体集中训练流程

3.2.1 预生成经验集

本文中采用预生成经验集代替常规强化学习中动态生成经验集的方式：提前通过批量训练和仿真生成大量经验样本，并通过特定采样策略，从中筛选出高效的经验集用于后续训练。整体流程如图 5 所示，具体包括以下两个步骤。

1) 原始经验样本生成：首先，通过几次常规强化学习训练，让智能体自动决策，并通过电力系统仿真获取大量原始经验样本。

2) 经验样本采样：在原始样本基础上，根据设定的采样策略筛选样本，构成最终用于训练的经验集。不同的采样策略会影响经验集的结构和训练效果，相关讨论将在算例分析部分详细展开。

值得注意的是，由于奖励系数常常需要调优，样本记录时可以保存计算奖励所需的全部原始量，即可根据不同参数实时计算奖励。预生成经验池完成后，后续训练与调参都可基于该经验池进行，无需再次与环境交互调用电力系统暂态仿真。

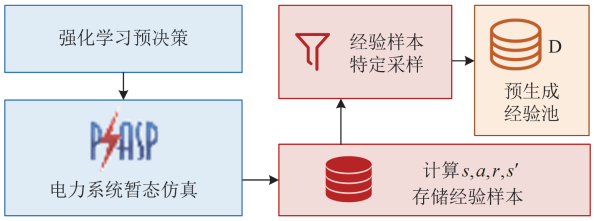


图 5 预生成经验池流程

Fig. 5 Pre-generated experience pool process

3.2.2 智能体集中训练

在完成预生成经验集后就可以开展智能体的集中训练。训练过程沿用强化学习时序差分损失计算与梯度下降参数更新方法，核心创新体现在训练模式优化。

常规强化学习方法采用小批量经验样本进行迭代更新，这种机制源于其在线数据收集与模型更新的同步特性。采用小批量样本训练，可以提升智能体对新收集经验样本的适应能力。但该策略导致样本利用效率低下，同时智能体的训练效果极易受样本质量波动影响，稳定性低。

本文采用预生成经验集驱动智能体的训练，在训练过程中不需要再收集新的经验样本。因此本文基于深度学习框架采用全部经验样本训练。每次训练过程中均采用经验池全部样本进行参数更新，从而显著提升样本利用效率。训练流程如图 6 所示。

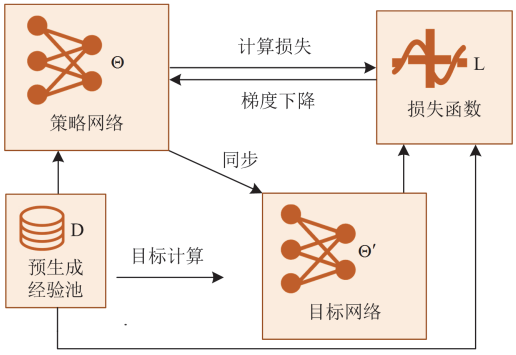


图 6 预生成经验集驱动的智能体训练流程

Fig. 6 A pre-generated experience set-driven process for training agent

4　算例分析

本文采用 10 机 39 节点系统验证所提算法的有效性。采用中国电科院研制的 PSASP 电力系统仿真计算软件进行紧急切负荷相关暂态过程的仿真。10 机 39 节点系统接线如图 7 所示。

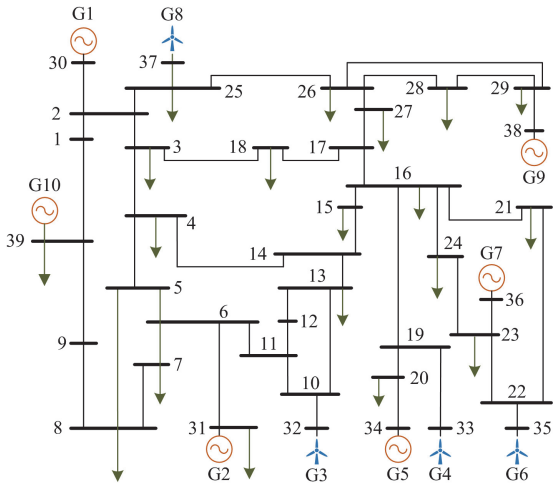


图 7　含新能源接入的 10 机 39 节点系统拓扑

Fig. 7 Topology of a 10-machine 39-node system with a high percentage of new energy access

系统 10 台发电机组分别采用同步发电机模型与风力发电机模型。整定 32、33、35、37 节点发电机为风力发电机，其余为同步发电机。风力发电机节点总容量为 2472 MW，系统总发电容量为 6140.81 MW，系统风电渗透率达 40.26%。

负荷模型采用静态负荷特性进行描述。其中，非紧急切负荷节点的负荷整定为带 50%恒定阻抗的恒功率模型，而紧急切负荷节点的负荷则采用带 0%恒定阻抗的恒功率模型。本文设定每个配置了紧急切负荷控制的节点，其负荷由 20 条出线平均分配，因此每次切除负荷实质上是切除该节点的部分负荷出线。基于此，每个节点的负荷切除情况可以表示为从 0%到 100%共 21 种状态。最终，整定母线 20、母线 24、母线 8 及母线 39 节点上的负荷为可切除负荷。

本文算法研究及验证硬件平台为 CPU Intel® Core(TM) i9-14900KF，RAM 128 GB，GPU Nvidia Geforce RTX 4080 Super，神经网络采用 Pytorch 库。

4.1　算例样本生成

电力系统频率崩溃通常由于系统突然失去大量有功出力导致。本文采用 PSASP 切机扰动模拟某时刻部分发电机突然脱网，导致系统有功出力骤降，进而引发功率缺额。研究在不同潮流水平下设置不同规模的发电机脱网情况，具体设置如表 1 所示。

定义故障发生于 0.1 s，紧急切负荷动作于故障发生后 0.1 s 投入。算例样本生成考虑发电机 $N-2$ 故障。依据参数设置表共得到 540 个预想故障样本。取 70%样本用于训练，剩余 30%样本用于评估决策结果。

表 1　故障样本生成条件

Table 1 Fault sample generation conditions

参数类型	参数具体设置	数量
故障种类	切机扰动	1
故障发电机组	全部 10 个发电机组	10
切机扰动大小	0.9, 1.0	2
系统潮流水平	发电机：100%　负荷：105%	3
	发电机：110%　负荷：115%	
	发电机：110%　负荷：125%	

4.2　预生成经验集与经验样本的特定采样

本文通过常规强化学习过程积累经验数据，共生成并利用 162 050 个经验样本。为评估经验样本的生成效率，进行了 10 次独立的强化学习训练，并统计每次训练中去重后的新增样本数量(见图 8)。结果表明，仅需 4~5 次训练即可满足本文算例对经验样本数量的需求。

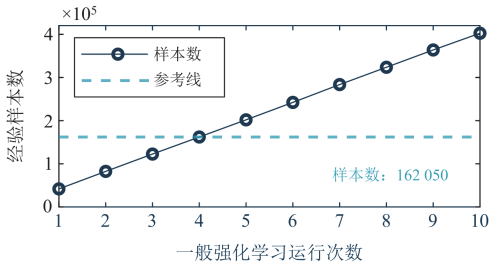


图 8　经验样本生成效率统计

Fig. 8 Empirical sample generation efficiency statistics

在完成经验样本的生成后，还需通过特定采样方法构建最终的经验集。这是因为常规强化学习训练过程中会产生大量低效决策样本，这些样本可能掩盖对训练具有关键引导作用的高价值经验，进而影响训练效率。

为提升训练效果，本文依据奖励值对经验样本进行采样，重点关注高奖励样本与低奖励样本。高奖励样本有助于引导智能体学习优良决策，低奖励样本则有助于智能体规避不良行为。在采样过程中，首先按设定比例从全部样本中选取高(或低)奖励样本，剩余部分则从未被选中的样本中随机补足，以保证采样后经验集的规模固定。最终，采集用于训练的经验样本约占全部生成样本的 70%。随后，针对高奖励采样、低奖励采样、不进行特定采样及不同采样比例等多种策略，分别对智能体训练效果进

行对比分析。图 9 展示了不同采样策略下的训练结果及经验集的样本分布。

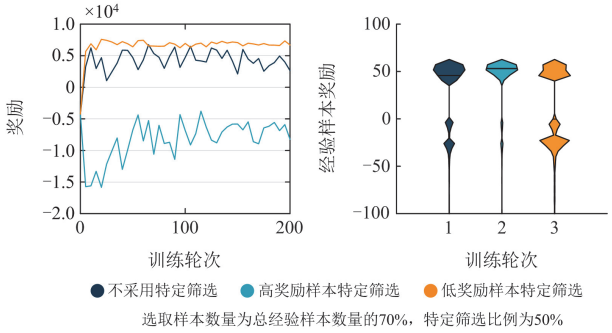


图 9　不同采样策略下训练效果对比

Fig. 9 Training performance comparison of different sampling strategies

实验结果表明，未进行特定采样时，智能体的训练效果较差。引入低奖励样本的特定采样后，训练效果显著提升；而采用高奖励样本特定采样则使训练结果进一步恶化。结合不同采样策略下的样本奖励分布(见小提琴图)可知，常规强化学习过程中生成的经验样本中低奖励样本占比较低，难以有效引导智能体规避不良决策，导致整体训练效果受限。高奖励样本采样进一步削弱了低奖励样本的占比，训练效果因此恶化。相比之下，通过低奖励样本特定采样，有效补充了经验集中低奖励样本的不足，有助于智能体学习规避无效或负面行为，从而提升了训练效果。

进一步分析低奖励样本采样比例对训练结果的影响，如图 10 所示。

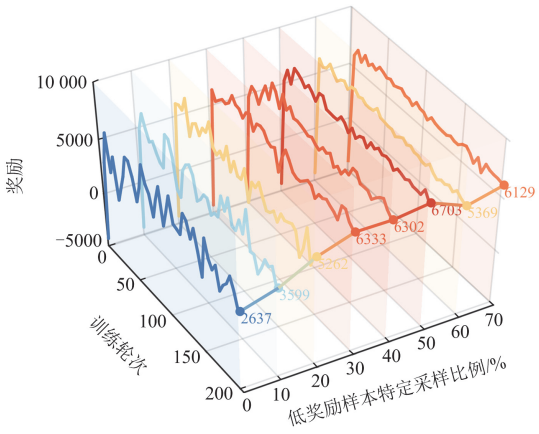


图 10　不同低奖励样本采样比例下的训练性能

Fig. 10 Training performance under different low-reward sampling ratios

实验结果显示，随着低奖励样本特定采样比例的增加，训练奖励值先升后降，在采样比例为 50%时达到最大值。原因在于，适度引入低奖励样本有助于智能体规避错误决策，但比例过高则导致正向经验样本不足，从而影响策略优化。因此，后续实验均采用 50%的低奖励样本采样比例。

4.3　智能体训练过程分析

基于智能体的训练过程，分析本文所提算法相比一般强化学习在智能体训练上的优势。两种算法下智能体训练的参数比较如表 2 所示。给出两种算法训练过程中智能体评估环节的奖励以及决策成功率曲线如图 11 所示。

表 2 智能体训练参数

Table 2 Agent training parameters

参数类型	本文算法	一般强化学习
总训练轮次	200	50 000
评估频率	5	2000
决策成功奖励 $r_{success}$	62	62
切负荷惩罚系数 λ_1	0.015	0.015
决策成功奖励系数 λ_2	22	22

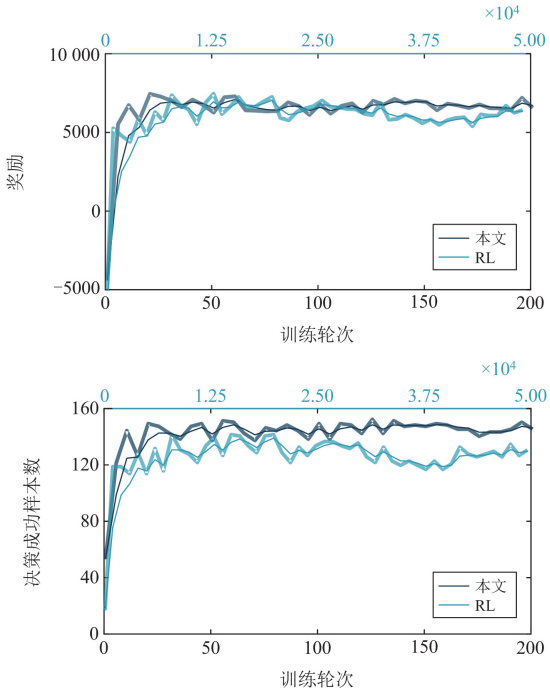


图 11　智能体训练评估曲线

Fig. 11 Agent training assessment curves

10 机 39 节点系统训练结果表明，相较于传统强化学习方法，本文算法展现出更优的收敛特性(约 200 轮次即达稳定状态)和更高的决策成功率。

为进一步比较智能体的训练时间成本，本文对两种算法分别进行了相同轮次的训练，并对训练用时进行了统计。在 10 机 39 节点系统规模下，传统强化学习方法一次训练耗时为 8 h 16 min 59 s，而本

文所提出的算法仅需 43 min 51 s，训练时间仅为前者的 8.82%。

4.4 智能体决策效果验证

4.4.1 决策效果评价指标

对于智能体决策效果的评价从 3 个方面进行，分别是决策的有效性、精确性与快速性。

1) 决策有效性。衡量切负荷决策是否有效主要关注决策后系统频率是否能保持稳定。若决策后系统频率维持在稳定范畴，则判定决策成功。

本文中，紧急切负荷后电力系统频率稳定的判据是系统最终恢复频率 $f_{\text{COI·stable}}$ 不低于 49.5 Hz 也不高于 50.5 Hz。暂态过程中电网各节点的频率会略有差异，因此采用系统惯性中心频率 f_{COI} 作为系统频率的衡量指标[30]。系统惯性中心频率是各发电机节点频率依照其惯量的加权平均，其计算表达式为

$$f_{\text{COI}}=\frac{\sum_{i=1}^{n_{\text{G}}}(H_i\cdot f_i)}{\sum_{i=1}^{n_{\text{G}}}H_i} \tag{7}$$

式中：H_i 是第 i 台发电机的惯性常数，通常以秒为单位；f_i 是第 i 台发电机的瞬时频率。相应的给出系统频率稳定的判据为

$$49.5\ \text{Hz}<f_{\text{COI·stable}}<50.5\ \text{Hz} \tag{8}$$

本文中采用紧急切负荷决策的成功率作为评判有效性的指标，如式(8)所示。

$$R_{\text{success}}=\frac{N_{\text{success}}}{N} \tag{9}$$

式中：R_{success} 为决策成功率；N_{success} 为决策成功样本数量；N 为决策样本总数。

2) 决策的精确性。精确切负荷要求在满足频率稳定的基础上尽可能少切除负荷。采用各决策成功样本的平均切负荷量作为决策精确性的参照。

$$L_s^{\text{avg}}=\frac{\sum_{i=1}^{N_{\text{success}}}(L_s^i)}{N_{\text{success}}} \tag{10}$$

式中：L_s^{avg} 为决策成功样本的平均切负荷量。

实际决策过程中，考虑到各样本所需切负荷量存在差异，不同算法的决策偏重也不尽相同，仅采用切负荷量难以全面衡量决策的精确性。系统的恢复频率与切负荷量息息相关，若算法进行不必要的过切负荷时，便会导致系统恢复频率升高。因此引入平均恢复频率偏差 Δf_{avg} 作为决策准确性的衡量指标。平均恢复频率偏差定义为各样本恢复频率与系统频率稳定的最低恢复频率(本文中是 49.5 Hz)之差，如式(11)所示。

$$\Delta f_{\text{avg}}=\frac{1}{N_{\text{success}}}\sum_{i=1}^{N_{\text{success}}}(f_{\text{COI·stable}}^i-49.5) \tag{11}$$

式中：$f_{\text{COI·stable}}^i$ 表示第 i 个样本的恢复频率。

3) 决策的快速性。决策的快速性是指算法对某一故障样本做出决策所需的时间。本文中通过统计各决策样本的平均决策时间来比较决策的快速性。

$$t_d^{\text{avg}}=\frac{t_d^{\text{total}}}{N} \tag{12}$$

式中：t_d^{avg} 表示平均决策时间；t_d^{total} 表示各决策样本总时间。

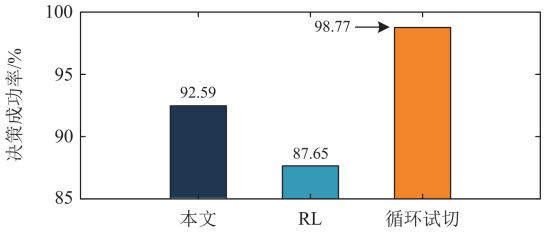
4.4.2 比较算法

比较算法 1：一般强化学习算法。其马尔可夫过程、智能体网络设计等均按照本文第 2 节所述，与本文所提算法保持一致。该算法采用常规的强化学习训练方式，即通过智能体与环境的实时交互获取经验，并同步进行网络训练与学习。

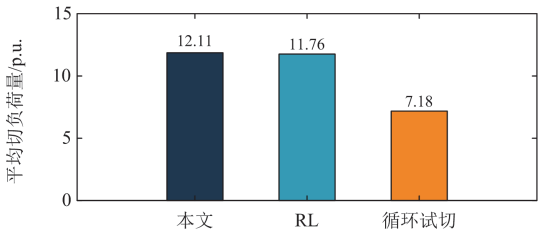
比较算法 2：循环试切负荷算法。文献[14]针对紧急切机问题，提出了一种基于循环仿真与灵敏度计算以实现紧急切机决策离线自动化制定的方法。该方法基于贪心策略，采用近乎遍历的方法得到切机决策，因此在决策的有效性与精确度上表现优异，可以视为决策的近似最优解。本文中将上述算法应用于紧急切负荷问题，作为切负荷决策的近似最优标准来衡量所提算法的效果。

4.4.3 各算法决策效果统计

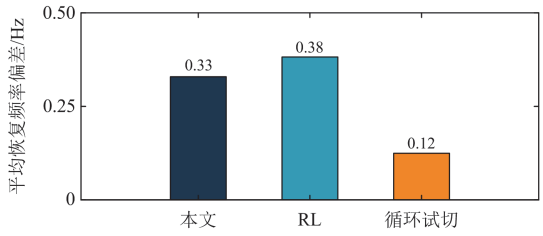
上述两种比较算法及本文所提算法在 10 机 39 节点系统 162 个评估样本上的决策效果如图 12 所示。



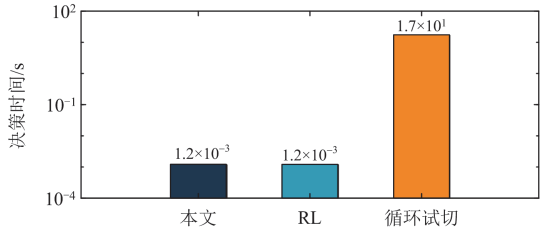
(a) 决策成功率



(b) 决策平均切负荷量



(c) 决策后平均恢复频率偏差



(d) 各算法单次决策时间

图 12 智能体决策效果统计

Fig. 12 Statistics on the effects of agent decision-making

1) 决策有效性比较。循环试切算法依赖于多次离线迭代试切，因此决策成功率较高，达到 98.77%。相比之下，在线决策场景下无法进一步仿真修正决策，因此成功率低于可离线多次修正的循环试切算法。然而同样在线决策下，相较于一般强化学习方法，本文所提算法决策成功率提升达 4.94%。

2) 决策精确性比较。离线的循环试切算法决策精确度最高。在线决策场景下，本文所提算法相比一般强化学习方法平均恢复频率偏差降低 0.05 Hz，决策精确度更高；平均切负荷量增加 0.35. p.u.，结合高决策成功率与低恢复频率偏差的分析，本文算法对于 10 机 39 节点系统中的严重故障适应性更好。

3) 决策快速性比较。针对离线场景的循环试切算法依赖多次试切迭代，决策时间极长，难以适应在线决策要求。此外，试切决策仅能针对当前故障工况，对复杂工况的适应性较差。相比之下，本文算法及一般强化学习方法在训练完成后，面对实时工况，仅需简单网络计算即可做出决策。

进一步给出各决策成功样本平均恢复频率偏差的分布如图 13 所示。

由系统平均恢复频率偏差分布直方图可以看出，相比一般强化学习，本文所提算法在决策结果上分布更为集中，整体恢复频率偏差更小，较大频率偏差样本更少。

10 机 39 节点系统算例表明，循环试切算法基于多次迭代试切保证了高成功率与高精确度，但由于决策时间较长，仅适用于离线场景。在在线决策场景下，本文算法在决策成功率与精确度上均优于

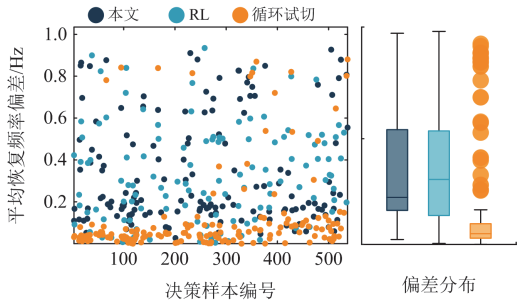


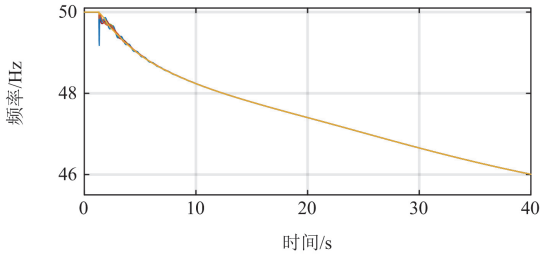
图 13 各算法平均恢复频率偏差分布

Fig. 13 Distribution of mean recovery frequency deviation across algorithms

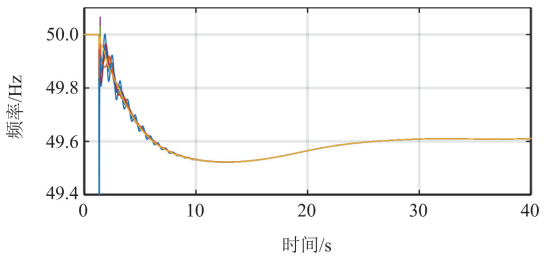
一般强化学习方法，并且已经接近离线的循环试切算法。这主要得益于基于预生成经验集的集中训练框架，显著降低了智能体的训练成本，使得更充分的调参成为可能。

4.4.4 切负荷决策案例分析

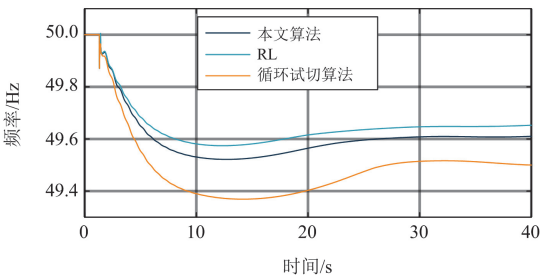
某给定工况下，各算法切负荷决策过程如图 14 所示。



(a) 未切负荷时各机组频率



(b) 本文算法切负荷决策后各机组频率



(c) 各算法切负荷后系统惯性中心频率

图 14 切负荷前后系统频率变化

Fig. 14 System frequency change before and after load cutting

未切除负荷时，系统频率失稳。本文算法训练智能体实时决策，共计切除 8.16 p.u.负荷后，系统恢复频率为 49.61 Hz，通过紧急切除负荷成功避免了系统发生频率失稳。比较各算法切负荷后系统惯性中心频率，在 10 机 39 节点系统下，本文算法的决策精确度优于一般强化学习方法，接近离线的循环试切算法。

5 结论

本文针对电力系统频率稳定问题，提出了一种基于预生成经验集驱动的频率稳定紧急切负荷智能在线决策方法。该方法结合频率稳定与在线决策的需求，基于马尔可夫决策过程构建了切负荷决策模型；通过引入 BDQ 算法，有效解决了离散负荷高维动作空间带来的决策难题。在智能体训练方面，采用预生成经验池驱动的方法，显著降低了训练时间和仿真成本，提高了智能体在超参数密集调节场景下的训练效率。

10 机 39 节点系统的算例结果表明，所提算法在该规模系统下的智能体训练时间仅为传统强化学习方法的 8.82%，显著降低了训练成本。在决策性能方面，所提方法的切负荷决策成功率较传统强化学习方法提升了 4.94%，同时平均频率恢复偏差减少了 0.05 Hz。与离线试切算法相比，所提算法在决策性能上表现接近，但决策时间大幅缩短，充分体现了其在在线决策场景下的高效性与实用性。

本文提出的算法在 10 机 39 节点系统中展现出了良好的训练效率和决策性能，充分验证了所提方法的有效性。针对更大规模的实际电网中更加复杂的响应机制和高维决策空间，相关的应用与推广还需要进一步深入研究。通过预生成经验池驱动训练，所提方法有效减少了对电力系统暂态仿真的频繁调用，在应对大电网高仿真成本方面展现出一定的应用潜力。未来的研究将重点关注该方法在大规模电网中的适用性验证，并融合电力系统领域知识，持续完善和优化算法，以应对大电网带来的新挑战。

参考文献

[1] 李东东, 孙雅茹, 徐波, 等. 考虑频率稳定的新能源高渗透率电力系统最小惯量与一次调频容量评估方法[J]. 电力系统保护与控制, 2021, 49(23): 54-61.
LI Dongdong, SUN Yaru, XU Bo, et al. Minimum inertia and primary frequency capacity assessment for a new energy high permeability power system considering frequency stability[J]. Power System Protection and Control, 2021, 49(23): 54-61.

[2] 汤明润, 李若旸, 程晓钰, 等. 基于博弈论-改进云模型的新能源电力系统适应性评估[J]. 电力系统保护与控制, 2025, 53(7): 40-51.
TANG Mingrun, LI Ruoyang, CHENG Xiaoyu, et al. Adaptability evaluation of new energy power systems based on game theory and an improved cloud model[J]. Power System Protection and Control, 2025, 53(7): 40-51.

[3] 钱敏慧, 张建胜, 秦文萍, 等. 计及 DFIG 调频的系统频率响应特性分析及快速频率支撑策略研究[J]. 电力系统保护与控制, 2025, 53(3): 58-67.
QIAN Minhui, ZHANG Jiansheng, QIN Wenping, et al. System frequency response characteristic considering DFIG frequency regulation and fast frequency response strategy[J]. Power System Protection and Control, 2025, 53(3): 58-67.

[4] 魏玖明, 李兆伟, 李碧君, 等. 提升新能源短时性功率冲击下暂态频率安全的储能紧急控制策略[J]. 电力系统自动化, 2024, 48(8): 152-161.
WEI Jiuming, LI Zhaowei, LI Bijun, et al. Emergency control strategies using energy storage to enhance transient frequency safety under short-time power impact of renewable energy[J]. Automation of Electric Power Systems, 2024, 48(8): 152-161.

[5] 刘彦伶, 武志刚, 赖翔, 等. 异质空调负荷参与多区域电力系统频率调节的协同控制策略[J]. 电力系统保护与控制, 2025, 53(3): 47-57.
LIU Yanling, WU Zhigang, LAI Xiang, et al. Cooperative control strategy for heterogeneous air-conditioning loads participating in frequency regulation of multi-area power systems[J]. Power System Protection and Control, 2025, 53(3): 47-57.

[6] 胡加伟, 王彤, 王增平. 直流闭锁后系统暂态稳定紧急协同控制策略研究[J]. 电力系统保护与控制, 2023, 51(4): 43-52.
HU Jiawei, WANG Tong, WANG Zengping. Collaborative emergency control strategy of system transient stability after DC blocking[J]. Power System Protection and Control, 2023, 51(4): 43-52.

[7] 李祖明, 张星宇, 陈松林, 等. 安稳系统中风电脱网风险在线评价指标[J]. 电力系统保护与控制, 2020, 48(20): 162-169.
LI Zuming, ZHANG Xingyu, CHEN Songlin, et al. On-line evaluation method of wind farm off-grid risk in power system stability control[J]. Power System Protection and Control, 2020, 48(20): 162-169.

[8] 胡泽, 曾令康, 姚伟, 等. 电力系统两阶段紧急切负荷控制智能预决策[J]. 中国电机工程学报, 2024, 44(4): 1260-1272.
HU Ze, ZENG Lingkang, YAO Wei, et al. Intelligent pre-decision of two-stage emergency load shedding control in power systems[J]. Proceedings of the CSEE, 2024, 44(4): 1260-1272.

[9] 潘晓杰, 胡泽, 姚伟, 等. 融合电网拓扑信息的分支竞争 Q 网络智能体紧急切负荷决策[J]. 电力系统保护与

控制, 2025, 53(8): 71-80.
PAN Xiaojie, HU Ze, YAO Wei, et al. Emergency load shedding decision-making using a branching dueling Q-network integrating grid topology information[J]. Power System Protection and Control, 2025, 53(8): 71-80.

[10] LI Q, XU Y, REN C. A hierarchical data-driven method for event-based load shedding against fault-induced delayed voltage recovery in power systems[J]. IEEE Transactions on Industrial Informatics, 2020, 17(1): 699-709.

[11] 续昕, 张恒旭, 李常刚, 等. 基于轨迹灵敏度的紧急切负荷优化算法[J]. 电力系统自动化, 2016, 40(18): 143-148.
XU Xin, ZHANG Hengxu, LI Changgang, et al. Emergency load shedding optimization algorithm based on trajectory sensitivity[J]. Automation of Electric Power Systems, 2016, 40(18): 143-148.

[12] 孙大雁, 周海强, 熊浩清, 等. 基于灵敏度分析的直流受端系统紧急切负荷控制优化方法[J]. 中国电机工程学报, 2018, 38 (24): 7267-7275, 7453.
SUN Dayan, ZHOU Haiqiang, XIONG Haoqing, et al. A sensitivities analysis based emergency load shedding optimization method for the HVDC receiving end system[J]. Proceedings of the CSEE, 2018, 38(24): 7267-7275, 7453.

[13] 强子玥. 基于深度学习的电力系统暂态失稳紧急控制策略[D]. 北京: 北京交通大学, 2021.

[14] 李舟平. 数据—知识融合驱动的电力系统智能紧急切机决策及原型系统设计[D]. 武汉: 华中科技大学, 2022.

[15] HU Z, YAO W, SHI Z, et al. Knowledge-enhanced deep reinforcement learning for intelligent event-based load shedding[J]. International Journal of Electrical Power & Energy Systems, 2023, 148.

[16] 文云峰, 杨伟峰, 林晓煌. 低惯量电力系统频率稳定分析与控制研究综述及展望[J]. 电力自动化设备, 2020, 40(9): 211-222.
WEN Yunfeng, YANG Weifeng, LIN Xiaohuang. Review and prospect of frequency stability analysis and control of low-inertia power systems[J]. Electric Power Automation Equipment, 2020, 40(9): 211-222.

[17] MEVLUDIN G. (Deep) Reinforcement learning for electric power system control and related problems: a short review and perspectives[J]. Annual Reviews in Control, 2019, 4822-4835.

[18] CAO D, HU W, ZHAO J, et al. Reinforcement learning and its applications in modern power and energy systems: a review[J]. Journal of Modern Power Systems and Clean Energy, 2020, 8(6): 1029-1042.

[19] HU Z, YAO W, SHI Z, et al. Intelligent and rapid event-based load shedding pre-determination for large-scale power systems: knowledge-enhanced parallel branching dueling Q-network approach[J]. Applied Energy, 2023, 347.

[20] HU Z, SHI Z, ZENG L, et al. Knowledge-enhanced deep reinforcement learning for intelligent event-based load shedding[J]. International Journal of Electrical Power & Energy Systems, 2023, 148.

[21] HUANG R, CHEN Y, YIN T, et al. Accelerated derivative-free deep reinforcement learning for large-scale grid emergency voltage control[J]. IEEE Transactions on Power Systems, 2021, 37(1): 14-25.

[22] 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.

[23] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing Atari with deep reinforcement learning[J]. arXiv preprint arXiv: 1312. 5602, 2013.

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

[25] 张峻伟, 吕帅, 张正昊, 等. 基于样本效率优化的深度强化学习方法综述[J]. 软件学报, 2022, 33(11): 4217-4238.
ZHANG Junwei, LÜ Shuai, ZHANG Zhenghao, et al. Survey on deep reinforcement learning methods based on sample efficiency optimization[J]. Journal of Software, 2022, 33(11): 4217-4238.

[26] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[M]. Cambridge: MIT Press, 1998.

[27] 张启阳, 陈希亮, 曹雷, 等. 深度强化学习中的知识迁移方法研究综述[J]. 计算机科学, 2023, 50(5): 201-216.
ZHANG Qiyang, CHEN Xiliang, CAO Lei, et al. Survey on knowledge transfer method in deep reinforcement learning[J]. Computer Science, 2023, 50(5): 201-216.

[28] TAVAKOLI A, PARDO F, KORMUSHEV P. Action branching architectures for deep reinforcement learning[C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1).

[29] KANERVISTO A, SCHELLER C, HAUTAMÄKI V. Action space shaping in deep reinforcement learning[C] // 2020 IEEE Conference on Games (CoG), August 24-27, 2020, Osaka, Japan: 479-486.

[30] ANDERSON P M, MIRHEYDAR M. A low-order system frequency response model[J]. IEEE Transactions on Power Systems, 1990, 5(3): 720-729.

收稿日期：2025-04-22；　修回日期：2025-07-29

作者简介：

李成翔(1994—)，男，硕士，工程师，研究方向为交直流大电网实时仿真、新能源并网仿真与控制；

龚楷程(2002—)，男，通信作者，硕士研究生，研究方向为人工智能在电力系统稳定分析与控制中的应用；E-mail: gongkaicheng@hust.edu.cn

姚　伟(1983—)，男，教授，博士生导师，研究方向为高比例新能源电力系统稳定分析与控制、新一代电力人工智能技术及应用。E-mail: w.yao@hust.edu.cn

(编辑　魏小丽)