DOI：10.13334/j.0258-8013.pcsee.222633 文章编号：0258-8013 (2024) 05-1675-13 中图分类号：TM 73 文献标识码：A

基于知识融合和深度强化学习的智能紧急切机决策

李舟平[1]，曾令康[1]，姚伟[1*]，胡泽[1]，帅航[2]，汤涌[3]，文劲宇[1]

(1. 强电磁工程与新技术国家重点实验室(华中科技大学电气与电子工程学院)，湖北省 武汉市 430074;
2. 田纳西大学电气工程与计算机科学系，美国 田纳西州 诺克斯维尔市 37996;
3. 中国电力科学研究院有限公司，北京市 海淀区 100192)

Intelligent Emergency Generator Rejection Schemes Based on Knowledge Fusion and Deep Reinforcement Learning

LI Zhouping[1], ZENG Lingkang[1], YAO Wei[1*], HU Ze[1], SHUAI Hang[2], TANG Yong[3], WEN Jinyu[1]

(1. State Key Laboratory of Advanced Electromagnetic Engineering and Technology (School of Electrical and Electronic Engineering, Huazhong University of Science and Technology), Wuhan 430074, Hubei Province, China;
2. Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville 37996, Tennessee, USA;
3. China Electric Power Research Institute, Haidian District, Beijing 100192, China)

ABSTRACT: Emergency control is an important means of maintaining power system transient security and stability following serious faults. The current popular "human-in-the-loop" offline emergency control decision-making method has some drawbacks, including low efficiency and heavy reliance on expert experience. Therefore, this paper proposes an intelligent emergency generator rejection decision-making method based on knowledge fusion and deep reinforcement learning (DRL). First, a DRL-based emergency generator rejection decision-making framework is built. Then, when the agent deals with multi-generator decisions, the resulting high-dimensional decision space makes the agent training difficult. There are two solutions proposed: decision space compression and the application of a branching dueling Q (BDQ) network. Next, to further improve the exploration efficiency and the decision-making quality of the agent, the knowledge and experience related to emergency generator rejection control are integrated to the agent training. Finally, the simulation results in the 10-machine 39-bus system show that the proposed method can quickly give effective emergency generator rejection decisions in multi-generator decision-making. Applying a BDQ network has better decision performance than decision space compression. The knowledge fusion strategy can guide the agents to reduce ineffective decision-making explorations and improve decision-making performance.

KEY WORDS: emergency generator rejection decision; deep reinforcement learning; decision space; branching dueling Q network; knowledge fusion

摘要：紧急控制是在严重故障后维持电力系统暂态安全稳定的重要手段。目前常用的“人在环路”离线紧急控制决策制定方式存在效率不高、严重依赖专家经验等问题，该文提出一种基于知识融合和深度强化学习(deep reinforcement learning，DRL)的智能紧急切机决策制定方法。首先，构建基于 DRL 的紧急切机决策制定框架。然后，在智能体处理多个发电机决策时，由于产生的高维决策空间使得智能体训练困难，提出决策空间压缩和应用分支竞争 Q(branching dueling Q，BDQ)网络的两种解决方法。接着，为了进一步提高智能体的探索效率和决策质量，在智能体训练中融合紧急切机控制相关知识经验。最后，在 10 机 39 节点系统中的仿真结果表明，所提方法可以在多发电机决策时快速给出有效的紧急切机决策，应用 BDQ 网络比决策空间压缩的决策性能更好，知识融合策略可引导智能体减少无效决策探索从而提升决策性能。

关键词：紧急切机决策；深度强化学习；决策空间；分支竞争 Q 网络；知识融合

0 引言

紧急控制是在发生严重故障后为保证电力系统安全稳定采取的重要控制措施[1]，一般有 3 种控

基金项目：国家自然科学基金项目(U1866602)。
Project Supported by National Natural Science Foundation of China (U1866602).

制模式：离线整定-实时匹配、在线整定-实时匹配和实时匹配-实时控制[2]。离线整定-实时匹配在失稳预想故障工况下整定出有效决策，在系统运行时根据实时监测状态匹配出决策并执行。该模式控制措施投入速度快，当实际故障能匹配预想故障时决策效果好[3]。而后两种模式利用实时量测信息快速计算决策，对于工况变化情况的匹配有一定适应性，但所需的超实时仿真和高效计算的仿真模型要求较高[4-6]。离线决策-实时匹配模式下决策经过仿真校核，是比较安全可靠的，因此目前实际电网中紧急控制主要采用该模式。

在传统的“人在环路”决策制定方式下，专家根据经验反复试凑出有效决策[7]。随着电网规模扩大、网架结构更复杂，这种手工决策方式存在一定局限性[8]。首先，运行方式繁多、故障形态复杂，决策制定高度依赖于特定系统的运行经验[9]。其次，海量数据分析的工作量耗费人力成本，效率不高。另外，专家的分析精度有限，决策并非最优。

为提升紧急切机决策的制定效率，已有文献中开展了一些决策制定方法的研究，主要可分为以下3类：1）相轨迹法[4,10]：文献[4]用相轨迹斜率构建不稳定性指标，识别关键发电机，通过等值计算求得决策近似解；2）暂态能量法[6,11]：文献[11]基于暂态能量分别在稳定和不稳定情形下计算裕度，然后插值法快速求解切机量。但这两类方法需要对系统进行等值计算，不适合用于含复杂动态特性的系统，而且也只考虑了决策量，忽略了决策对象的选择；3）启发式算法[12-14]：文献[12]将问题转化为带暂稳约束的优化控制来求解最优决策量，但应用中可能存在收敛性问题，在控制对象多时搜索空间较大，计算量也不小。

近年来，深度强化学习(deep reinforcement learning，DRL)[15-16]算法快速发展，已广泛应用于电力系统稳定控制[17-20]的相关研究中，测试结果充分体现了 DRL 智能体具有强大的特征提取和决策改进的能力。在紧急控制方面，文献[18]提出基于深度 Q 网络的紧急切机决策制定算法，利用随机矩阵理论计算的奖励训练智能体给出合适决策。文献[19]提出基于 DRL 的紧急控制决策制定框架，在发电机紧急制动控制和低压减载控制场景下进行验证。但现有基于 DRL 的紧急控制算法应用于离线切机决策制定时存在如下问题：

1）不适合离线决策制定场景：目前文献中 DRL 实现紧急切机一般是针对实时闭环控制场景，智能体的马尔可夫决策过程(Markov decision process，MDP)的每个时间步长都要作出决策[19]，一次暂稳仿真对应多次决策。但对于本文处理的离线决策制定场景，智能体只在故障后作一次决策，每次尝试后都要进行一次完整的暂稳仿真校验其有效性，智能体可能需要尝试多次最终才能得到有效决策。因此，现有 DRL 智能体的 MDP 并不直接适用，需要重新设计。

2）不便于处理多发电机决策：当系统运行状态复杂或故障严重时，需要多个发电机参与切机决策才能使系统恢复稳定。文献[18]所提方法验证了单一发电机决策的情况，但如果直接用在多发电机决策场景容易出现决策空间维数爆炸问题，难以直接训练；文献[20]提出用多智能体算法分别处理多个发电机的决策，但协调多智能体训练难度较大。

3）纯数据驱动训练效率低：智能体采用常规的随机试错方式时，很多决策尝试明显没有学习价值，训练效率不高，而融合电力系统的知识经验可以减少其无效决策探索，提升训练效率[21]。目前，应用知识融合的研究较少，有必要探索领域知识和 DRL 的有机结合。

针对上述问题，本文提出基于知识融合和 DRL 的智能紧急切机决策方法。首先，根据紧急切机决策制定问题构建智能决策框架，针对当前 DRL 算法的 MDP 不适应离线决策制定场景的问题，设计智能体的交互决策过程；然后，针对智能体不便于多发电机参与决策的问题，设计智能体动作与决策空间；进一步，针对当前纯数据驱动方法的智能体训练效率低的问题，提出一些基于知识融合的改进策略；最后，在 10 机 39 节点系统的仿真结果验证本文所提方法的有效性。

1 基于 DRL 的紧急切机决策制定

1.1 紧急切机决策制定

图 1 为紧急切机离线整定-实时匹配的流程。在离线整定阶段，首先，专家根据经验确定可能引起功角失稳的预想故障工况集合$\{F^{(1)}, F^{(2)}, \cdots, F^{(M)}\}$。然后第 j 次抽取预想故障 $F^{(j)}$结合初始空决策 $u_{\mathrm{G}}^{(j)}$构成仿真条件$(F^{(j)}, u_{\mathrm{G}}^{(j)})$，执行暂态仿真计算，根据功角$\delta$判断功角稳定性。如果功角稳定，则考察第$j$+1 个预想故障，否则根据经验结合仿真数据分析出可能最有效的决策$u_{\mathrm{G}}'^{(j)}$，然后再次校核。如果决

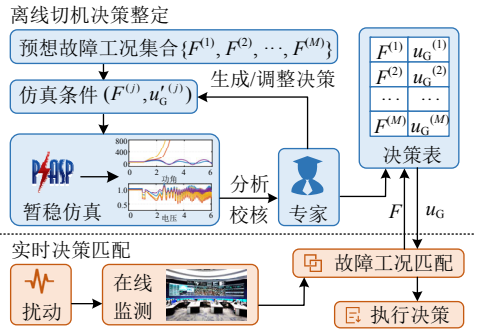


图 1　电力系统紧急切机决策的离线整定和实时匹配

Fig. 1　Offline determination and real-time match for emergency generator rejection decisions of power systems

策 $u_G'^{(j)}$ 有效则将其添加到决策表中，否则需要继续分析仿真数据并尝试新决策。若多次尝试也无法找到有效决策，则该潮流工况可能运行风险较大，需要调整潮流直到能满足系统安全稳定运行的要求[22]。

在实时匹配阶段，系统在受到扰动后，根据在线监测数据从决策表中匹配出对应故障工况及决策，然后执行决策保证系统安全稳定运行。

1.2　智能紧急切机决策制定框架

基于 DRL 的智能紧急切机决策制定的思路是用智能体替代图 1 中的专家。图 2 为基于知识融合 DRL 的智能紧急切机决策制定的通用框架。框架中左边是需要进行切机决策制定的电力系统环境，右边的 DRL 智能体可以使用深度 Q 网络(deep Q network，DQN)族算法，两边通过输入特征预处理的特征-状态转换($f\rightarrow s$)和将动作解码为实际切机决策的动作-决策转换($u\leftarrow a$)实现对接。该框架支持融合一些领域知识经验，如可以根据经验引入动作屏蔽 X 限制决策探索空间，也可以根据经验预先获得优质经验池 D_c，然后混合抽样进行智能体的训练。

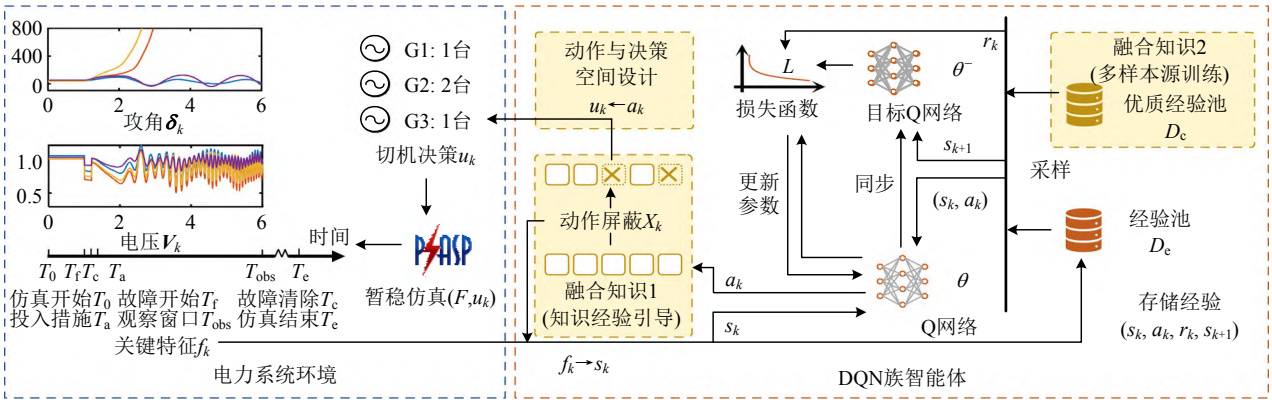


图 2　基于知识融合和深度强化学习的电力系统智能紧急切机决策制定框架

Fig. 2　Framework of intelligent emergency generators rejection decision for power system based on deep reinforcement learning with knowledge fusion

每个回合中，智能体根据输入状态尝试给出决策并进行一次时域仿真，如果系统功角稳定则回合结束，否则继续寻找决策。利用智能体与环境交互产生的样本训练智能体，最终的目标是使智能体对于失稳故障工况能通过尽量少的决策尝试得到切机代价尽量小的有效决策。

1.3　马尔可夫决策过程的设计

应用 DRL 需要将问题用 MDP 表达，下面介绍状态 s、奖励 r、动作 a 等 MDP 元素的设计：

1）特征 f_k 与状态 s_k：智能体的输入状态 s_k 要包含反映功角稳定和切机决策的关键信息[18,23]。本文选取观察时间窗口$[0,T_{obs}]$内的母线电压幅值矩阵 $\boldsymbol{V}_k$ 和发电机的相对功角矩阵$\boldsymbol{\delta}_k$ 作为特征 f_k。

$$\boldsymbol{V}_k=[V_1^k,\cdots,V_{n_B}^k]=\begin{bmatrix} V_{1,0}^k & \cdots & V_{n_B,0}^k \\ \vdots & \ddots & \vdots \\ V_{1,t_{obs}}^k & \cdots & V_{n_B,t_{obs}}^k \end{bmatrix} \tag{1}$$

$$\boldsymbol{\delta}_k=[\delta_1^k,\cdots,\delta_{n_G}^k]=\begin{bmatrix} \delta_{1,0}^k & \cdots & \delta_{n_G,0}^k \\ \vdots & \ddots & \vdots \\ \delta_{1,t_{obs}}^k & \cdots & \delta_{n_G,t_{obs}}^k \end{bmatrix} \tag{2}$$

式中：n_B 为监测电压幅值的母线数量；$V_{n,t}^k$ 为第 k 步下母线 n 采样时刻 t 的电压幅值；n_G 为监测功角发电机数量；$\delta_{n,t}^k$ 为第 k 步下发电机 n 采样时刻 t 的相对功角。为方便神经网络训练，对相对功角进行预处理，本文采用函数ϕ将功角值映射到[–1,1]区间。

$$\phi(\delta)=\tanh(\delta/2\pi) \tag{3}$$

最后，状态 s_k 可以用下式来表达：

$$s_k=(\boldsymbol{V}_k,\phi(\boldsymbol{\delta}_k)) \tag{4}$$

2）动作 a_k 与决策 u_k：决策 u_k 直接对应实际切机决策，动作 a_k 通过动作-决策转换与决策 u_k 对应，具体转换方式在 2 节中介绍。

$$\boldsymbol{u}_k=[u_{k(1)},u_{k(2)},\cdots,u_{k(n_G)}] \tag{5}$$

$$\boldsymbol{a}_k=[a_{k(1)},a_{k(2)},\cdots,a_{k(m)}] \tag{6}$$

式中：$u_{k(i)}$表示第 k 步下第 i 个节点的切机台数($0\le u_{k(i)}\le n_{Gi}$)，有 n_{Gi}+1 个切机选择；$a_{k(i)}$是第 k 步神经网络输出层第 i 个神经元的输出。

3）切机决策成功标识 p_k 与回合结束标识 e_k：回合结束有 2 种条件可以触发，一个是切机决策成功，即智能体给出的切机决策能使得系统恢复功角稳定；另一个是智能体达到预设的最大迭代次数 N_L，阻止智能体无穷尽搜索决策。故切机决策成功标识 p_k 与回合结束标识 e_k 可以表示如下。

$$p_k=1-S(\boldsymbol{\delta}_k;u_k) \tag{7}$$

$$e_k=\begin{cases}0, p_k=0\cap k\le N_L\\ 1, p_k=1\cup k>N_L\end{cases} \tag{8}$$

式中 S 为功角判稳函数，约定失稳为 1，稳定为 0。

4）奖励 r_k：奖励由两部分构成，一部分是对于切机决策成功的奖励，另一部分是根据实际切机量大小来确定的惩罚。

$$\begin{cases}r_k=\lambda_1\Delta U_k+\lambda_2 p_k\\ \Delta U_k=\sum\limits_{i=1}^{n_G}u_{k(i)}-\sum\limits_{i=1}^{n_G}u_{k-1(i)}\end{cases} \tag{9}$$

式中：ΔU_k 表示第 k 步与第 k–1 步之间切机量的差值；λ_1 为切机量惩罚系数；λ_2 为切机决策成功的奖励系数。通过奖励设置引导智能体尽可能在切机量小的情况下给出有效切机决策。

2 智能体的动作与决策空间设计

智能体对多个可控发电机节点进行决策时，每个动作 a 对应 N 个子动作的组合，每个子动作各有 n 个选择，则动作空间的维数为 n^N，对应的决策空间称为指数决策空间。若应用无分支的 DQN 算法，需要进行动作空间转换。但当 N 或者 n 较大时，动作空间维数太高，智能体的训练难度很大。由此可以考虑两种处理方法：决策空间压缩和应用分支竞争 Q (branching dueling Q，BDQ)网络。

2.1 决策空间压缩

决策空间压缩的处理方法是第 k+1 步智能体在第 k 步决策基础上选择一个节点进行调整，其他节点的决策不变，此时决策空间维度是 nN，与节点数成正比，可称为线性决策空间。

第 k 步时按照式(10)、(11)的ε-贪婪策略π^{ε}计算决策机组序号 x_k。

$$\begin{cases}x_k=\pi^{\varepsilon}(s_k,a_k,\varepsilon_k)\\ P(x_k=\arg\max\limits_x Q(s_k,a_k^x))=1-\varepsilon_k\\ P(x_k=x_{\text{random}})=\varepsilon_k\end{cases} \tag{10}$$

$$\varepsilon_k=\varepsilon_e+(\varepsilon_s-\varepsilon_e)e^{-n_s/n_d} \tag{11}$$

式中：ε_s 为起始贪婪因子；ε_e 为收敛贪婪因子；n_d 为贪婪因子衰减常数；n_s 为当前的训练迭代次数；x_{random} 为随机决策机组序号。

按式(12)可求解出切机节点序号 i_k、对应切机量大小 $v_{k(i_k)}$。

$$\begin{cases}x_k=n_{Gi}(i_k-1)+v_{k(i_k)}+1,\ 1\le x_k\le\sum\limits_{i=1}^{n_G}(n_{Gi}+1)\\ 1\le i_k\le n_G,\ 0\le v_{k(i_k)}\le n_{Gi}\end{cases} \tag{12}$$

第 k+1 步的决策 $u_{k(i_k)}$生成方式如下：

$$\begin{cases}u_{0(i)}\leftarrow 0,\ 0\le i\le n_G\\ u_{k+1}\leftarrow u_k,\ u_{k+1(i_{k+1})}\leftarrow v_{k+1(i_{k+1})}\end{cases} \tag{13}$$

这种决策方式是对智能体决策尝试次数的妥协，通过动作空间的转换可降低智能体训练难度。

2.2 应用 BDQ 网络

如图 3 所示，BDQ 网络[24]是一种结合双重 Q 网络[25]和竞争 Q 网络[26]的多优势分支变体。应用 BDQ 结构，可以直接处理多维动作空间。输入状态 s 经过神经网络输出共享表征 f，然后价值分支从表征 f 映射到价值 $V(s)$，N 个优势分支从表征 f 映射到 N 个子动作的优势 $A_d(s,a_d),\ d=1,2,\cdots,N$。分支 d 的动作 $a_d\in\mathcal{A}_d$，动作空间 $\mathcal{A}_d$ 的维数 $|\mathcal{A}_d|$ 为 n，Q 值 $Q_d(s,a)$按照竞争 Q 网络方式计算。

$$Q_d(s,a)=V(s)+\left[A_d(s,a_d)-\frac{1}{|\mathcal{A}_d|}\sum_{a'_d\in\mathcal{A}_d}A_d(s,a'_d)\right] \tag{14}$$

更新 Q 网络时分支 d 的目标 Q 值 Y_d，Y_d 可以按照双重 Q 网络的方式计算。

$$Y_d=r+\gamma Q_d(s',\arg\max_{a_d'\in\mathcal{A}_d}Q_d(s',a_d';\theta);\theta^-) \tag{15}$$

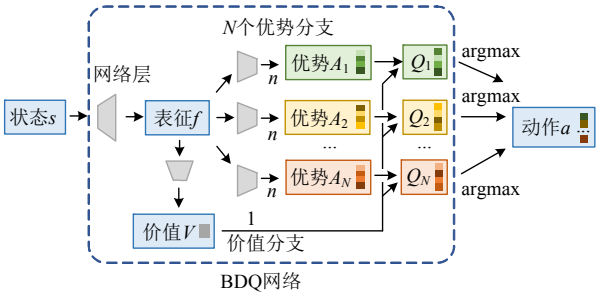


图 3　BDQ 网络的结构

Fig. 3　Structure of BDQ network

BDQ 的损失函数 L 定义为所有分支的 Q 值和目标 Q 值之间的均方误差。

$$L=\mathbb{E}[\frac{1}{N}\sum_{d}(Y_d-Q_d(s,a_d;\theta))^2] \tag{16}$$

如图 4 所示，每一行是神经网络的输出，对应着一个动作 a_k 和一个决策 u_k，使用常规无分支的 DQN 结构时，每次只有一个神经元被激活，用深底色表示一个发电机节点的一种决策量大小。下一次决策会保留了上一次决策，用浅底色标识，这样就可以实现对多个发电机节点的决策。而多分支的 BDQ 结构由于有多个输出分支，所以每次有多个神经元激活，每行有多个深底色标识。BDQ 结构可以直接处理指数决策空间，最理想情况下智能体可以训练到对所有预想故障工况都只用 1 步就给出有效决策。第 k 步决策 u_k 可表示为

$$\begin{cases} x_{k(i)}=\pi^{\varepsilon}(s_k,a_{k(i)},\varepsilon_k) \\ u_k=[u_{k(1)},u_{k(2)},\cdots,u_{k(n_G)}] \\ u_{k(i)}=v_{k(i)}=x_{k(i)}-1,\ 1\le i\le n_G \end{cases} \tag{17}$$

式中 $a_{k(i)}$为第 k 步第 i 个子动作神经网络的输出。

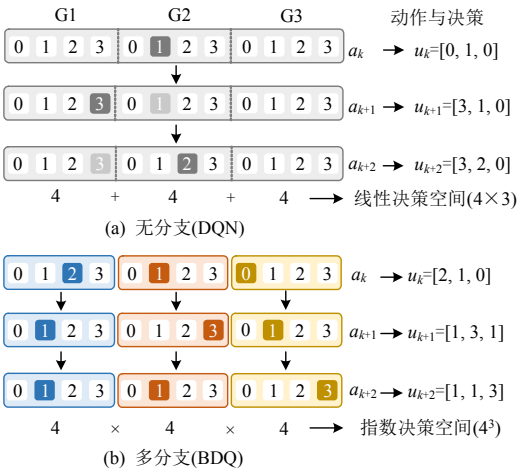


图 4　线性决策空间与指数决策空间的对比

Fig. 4　Comparison between linear decision space and exponential decision space

2.3　动作-决策转换关系

如 1.3 节所述，动作与决策之间需要进行转换才能实现智能体与环境的交互，其间的转换关系分为两种：覆盖式和增量式。2.2 节是覆盖式决策，动作 a_k 直接对应决策 u_k，决策 u_{k+1} 部分覆盖决策 u_k。而增量式决策中动作表示决策的增量，决策 u_{k+1} 由决策 u_k 叠加动作 a_{k+1} 得到。图 5 为两种决策方式的对比，以多分支结构 BDQ 的决策为例对增量式结构进行说明，无分支结构 DQN 可以类推得到。

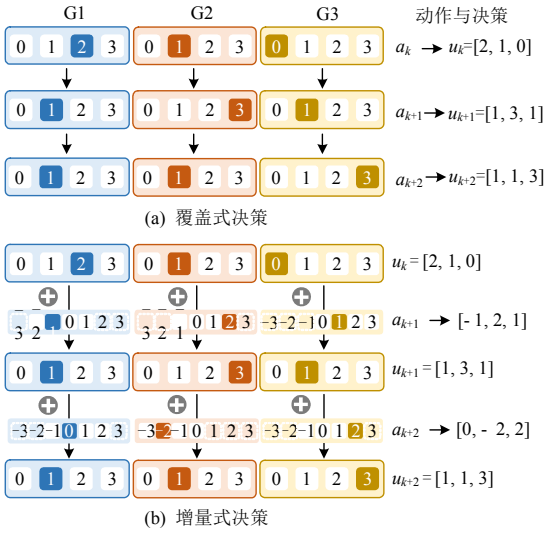


图 5　覆盖式决策与增量式决策的对比

Fig. 5　Comparison of overwritten decisions and incremental decisions

按照式(18)可以将动作 a_{k+1} 转换成决策机组序号 x_{k+1}、决策 u_k、u_{k+1} 和序号 x_{k+1}。

$$\begin{cases} x_{k+1(i)}=\pi^{\varepsilon}(s_{k+1},a_{k+1(i)},\varepsilon_{k+1}) \\ v_{k+1(i)}=x_{k+1(i)}-n_{Gi}-1 \\ u_{k+1(i)}=u_{k(i)}+v_{k+1(i)} \end{cases},\ 1\le i\le n_G \tag{18}$$

序号 x_{k+1} 有$-n_{Gi}$~n_{Gi}共 $2n_{Gi}$+1 个可选值，但决策 u_k 会对序号 x_{k+1} 的可选值有一定限制。这里引入屏蔽集合 X_{k+1}，X_{k+1} 由 n_G 个屏蔽子集 $X_{k+1(i)}$组成。

$$X_{k+1}=\{X_{k+1(1)},X_{k+1(2)},\cdots,X_{k+1(n_G)}\} \tag{19}$$

其中每个子集 $X_{k+1(i)}$中的元素 $x_{k+1(i)}$满足：

$$\begin{cases} u_{k+1(i)}=u_{k(i)}+x_{k+1(i)}-n_{Gi}-1 \\ u_{k+1(i)}<0 \ \text{或}\ u_{k+1(i)}>n_{Gi} \end{cases} \tag{20}$$

整理可得第 i 个节点的屏蔽子集 $X_{k+1(i)}$。

$$\begin{aligned} X_{k+1(i)}=\{x_{k+1(i)}\,|\,&x_{k+1(i)}<-u_{k(i)}+n_{Gi}+1 \\ &\text{或}\ x_{k+1(i)}>-u_{k(i)}+2n_{Gi}+1\} \end{aligned} \tag{21}$$

用式(22)屏蔽子集 $X_{k+1(i)}$修正序号 $x_{k+1(i)}$如下：

$$\begin{cases} x_{k+1(i)} = \pi^{\varepsilon}(s_{k+1}, a_{k+1(i)}, \varepsilon_{k+1}) \\ 1 \le i \le n_{\mathrm{G}},\ x_{k+1(i)} \notin X_{k+1(i)} \end{cases} \tag{22}$$

在强化学习中引入动作屏蔽可以避免智能体给出不符合实际情况的动作。使用增量式决策时需要将屏蔽集合 X_k 也作为状态输入给智能体。修改式(4)可得增量式决策下的状态 s_k。

$$s_k = (\boldsymbol{V}_k, \boldsymbol{\phi}(\boldsymbol{\delta}_k), X_k) \tag{23}$$

3 基于知识融合的智能体训练

知识融合是在数据驱动方法中融合相关知识引导智能体的训练，融合方式没有固定范式。本节介绍两种体应用知识融合的改进策略，一种是利用知识经验约束智能体的探索空间，另一种是混合优质决策样本进行智能体的训练。

3.1 应用知识经验约束探索

如图 6(a)所示，在前文介绍的常规的强化学习智能体训练方式中，智能体在决策探索时完全随机，没有切机节点和切除机组数量的限制，因而在实际训练时很容易出现决策无效或决策量偏大的情况，经验池的样本决策质量不够高。

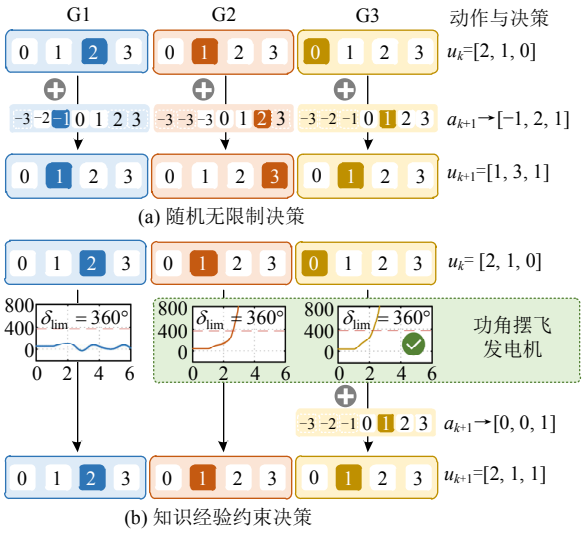


图 6 随机无限制决策与知识经验约束决策的对比

Fig. 6 Comparison between random unrestricted decision and knowledge experience constrained decision

人类专家在尝试决策时通常会用知识经验缩小决策范围，减少无效探索。因此，本文尝试借鉴人类智慧，将知识经验转化为智能体决策机组的选择范围的约束条件。根据经验，一般认为从功角摆飞的发电机节点中切机更有可能找到有效决策[27]，故本文对切机节点的选取进行限制。

切机节点限制可以通过在智能体输出动作之前加入动作屏蔽实现。如图 6(b)所示，每次只从功角摆飞切机节点集合 I^{u}_{k+1} 中随机选择一个节点 i^{s}_{k+1} 来切机，切除机组的数量也随机给定。屏蔽子集 $X_{k+1(i)}$可改写成式(25)。

$$I^{\mathrm{u}}_{k+1} = \{i \mid S(\delta_i^{k+1}) = 1, i = 1, 2, \cdots, n_{\mathrm{G}}\} \tag{24}$$

$$X_{k+1(i)} = \begin{cases} \{x_{k+1(i)} \mid x_{k+1(i)} < -u_{k(i)} + n_{\mathrm{G}i} + 1 \text{ 或} \\ x_{k+1(i)} > -u_{k(i)} + 2n_{\mathrm{G}i} + 1\}, & i = i^{s}_{k+1} \\ \{x_{k+1(i)} \mid x_{k+1(i)} \ne n_{\mathrm{G}i}\}, & i \ne i^{s}_{k+1} \end{cases} \tag{25}$$

在添加知识经验的限制后，智能体理论上可以在与环境的交互过程中产生更多有效探索的样本。在计算资源和时间有限的情况下，用这些样本训练智能体能获得更好的决策性能。

3.2 多样本源混合采样训练

如图 7(a)所示，在前面介绍的常规的强化学习智能体训练方式中，智能体与仿真环境交互产生样本并放入经验池中，然后进行经验采样来训练智能体。但训练开始时智能体是随机初始化的，因此初期智能体在交互中很难产生足够高质量的样本来学习。

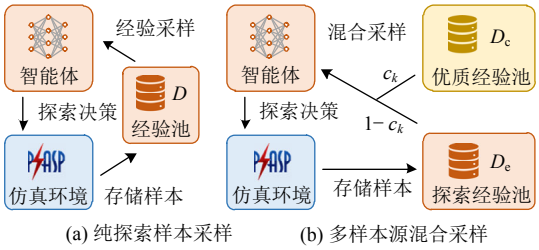


图 7 纯探索样本采样与多样本源混合采样的对比

Fig. 7 Comparison between pure exploratory sampling and mixed sampling of multiple sources

如果在初期引入一些高质量的决策训练智能体，有助于引导智能体学习优质决策的特点，从而提升决策性能。实际上，每年国家电网都会请大量专家进行紧急控制决策的制定。经验丰富的专家对于切机决策的机组选择等关键问题通常能很快给出高质量的决策方案[8]。这些高质量的决策就可以用来提升智能体训练的效率。

如图 7(b)所示，本文使用两个样本源混合训练智能体，一个样本源是纯探索的经验池 D_{e}，和图 7(a)中的常规训练方式一样，这个经验池中是智能体和环境交互所产生的探索样本。另一个样本源是优质经验池 D_{c}，其中是失稳预想故障工况训练集上的一些优质紧急切机决策样本。这些样本可以是专家根据知识经验手动制定得到，也可以是用优化

算法求得的优质决策样本。本文使用遍历计算求得近似最优切机决策样本。如式(26)所示，多样本源混合训练每次从 D_e 和 D_c 两个子经验池中按比例各自抽取部分样本组成大小为 B 的一批样本 $D(B)$，

$$\begin{cases} B_{c,k} = \lceil c_k B \rceil \\ B_{e,k} = B - B_{c,k} \\ D(B) \leftarrow \{D_e(B_{e,k}), D_c(B_{c,k})\} \end{cases} \tag{26}$$

$$c_k = c_e + (c_s - c_e)e^{-n_s/n_c} \tag{27}$$

式中：$B_{c,k}$ 和 c_k 分别表示第 k 步从 D_c 中抽取优质经验样本的数量和比例；$\lceil * \rceil$ 表示向上取整运算；$B_{e,k}$ 表示第 k 步从 D_e 中抽取的样本数量。

3.3　智能体的训练流程

智能体按照回合进行训练，每个回合从预想故障集中抽取失稳故障工况进行决策制定，每次交互产生的样本将存入经验池 D_e 中，每个回合结束后从池中抽取批大小 B 的一批样本，按照下式计算更新的智能体 Q 网络参数 θ。

$$\theta \leftarrow \theta + \alpha \nabla_\theta L \tag{28}$$

每隔 K 个回合同步一次目标 Q 网络参数 θ^-。智能体训练的完整流程参见图 8。

4　算例分析

4.1　环境与智能体参数设置

本文在 10 机 39 节点系统上进行算例验证，采用如表 1 所示的设置生成预想故障工况集合，其中包括 1 782 个故障工况。仿真经验发现，切机决策通常具有区域性，可以按两阶段决策方式进行切机决策。首先划分切机决策区域，确定可控切机决策节点，然后训练智能体确定每个工况的切机决策量。在所有工况中，由 G4—G7 这 4 个发电机节点构成的区域中暂态功角失稳工况共有 170 个，随机选择 130 个用于智能体的模型训练，剩余 40 个用于智能体的性能测试。

1）电力系统环境中的自动化交互：智能体与电力系统环境进行自动化交互是实现紧急切机决策自动化、智能化制定的基础。自动化交互要求电力系统环境能够根据每个预想故障工况仿真条件 (F, u_k)，自动调用仿真软件 PSASP 计算得到对应工况下智能体的输入特征 f_k。

为此，本文利用 Python 代码编程实现自动化仿真计算，其中 3 个主要功能如下：①配置文件修改功能：Python 平台按仿真条件 (F, u_k) 修改仿真时间、故障位置、决策发电机节点、输出变量列表等仿真软件 PSASP 使用的相关仿真配置文件。②电力仿

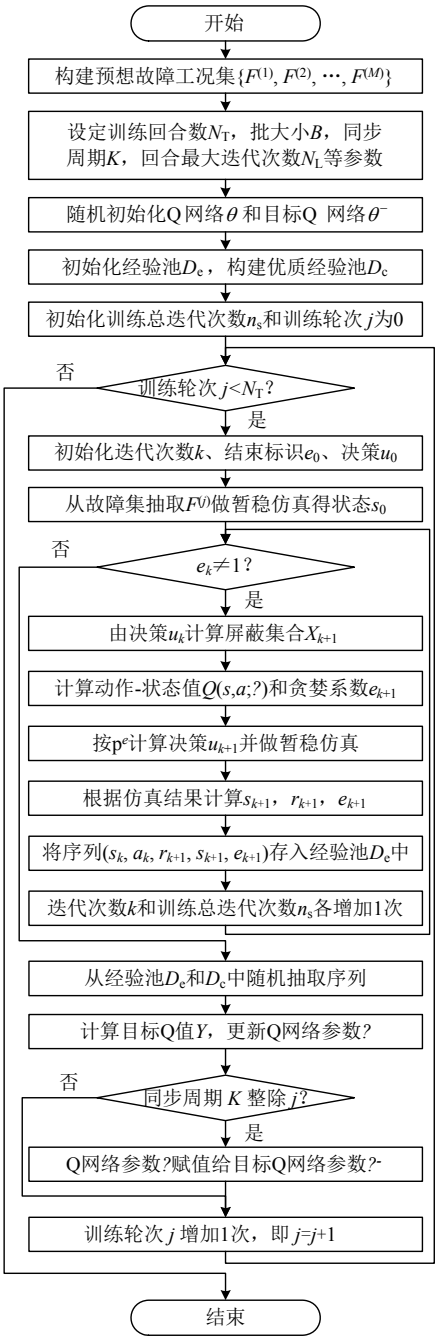


图 8　智能体训练流程图

Fig. 8　Flow chart of intelligent agent training

表 1　10 机 39 节点系统的预想故障工况设置

Table 1　Anticipated contingency settings for 10-machine 39-bus system

类型	具体设置	数量
故障类型	三相金属性短路故障	1
故障线路	除线路 16-19 外的所有交流线路	33
故障持续时间/s	0.2，0.25	2
故障位置/%	10，20，…，90	9
潮流水平	(1.0, 1.05)，(1.1, 1.15)，(1.2, 1.25)	3

真计算功能：Python 平台调用仿真软件 PSASP 的潮流计算、暂稳计算的可执行程序实现仿真计算。③提取仿真结果功能：Python 平台从仿真软件 PSASP 生成的仿真结果文件中按照预设的输出变量列表提取数据，进一步作数据预处理可得输入特征 f_k。

2）仿真参数设置与输入特征：系统的仿真步长设为 0.05 s，智能体输入观测窗口为[0, T_{obs}]，这里 T_{obs} 设为 6 s，一共有 121 个时间采样点，但是实际执行仿真的时长是 T_e=20 s。智能体输入包括两个张量，所有发电机的功角$\boldsymbol{\delta}$和所有母线的电压幅值$\boldsymbol{V}$，维度分别为(N,9,121)和(N,39,121)，N 表示批样本数。

3）神经网络的结构设计：如图 9 所示，智能体的神经网络分为特征提取部分和 Q 值输出部分。特征提取部分有两个一维卷积神经网络分支，每个分支的网络层包括卷积层、池化层和 ReLU 层。两个分支的输出展平后拼接作为 Q 值输出部分输入。

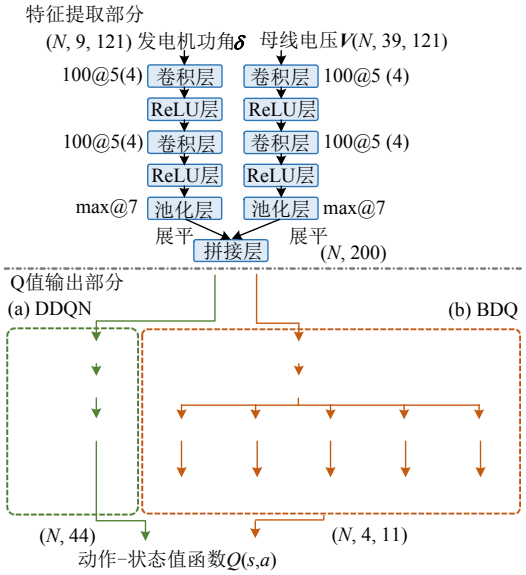


图 9 智能体的神经网络结构

Fig. 9 Neural network structures of the agent

Q 值输出部分可以采用 DDQN 网络或者 2.2 节介绍的 BDQ 网络来输出动作-状态值 $Q(s, a)$。根据 2.3 节中的介绍，智能体可以采用覆盖式和增量式两种动作-决策关系。两种网络结构和两种动作-决策关系进行组合可以形成覆盖式 DDQN、增量式 DDQN、覆盖式 BDQ、增量式 BDQ 4 种结构。

理论上，使用 BDQ 网络相比 DDQN 网络能更好地处理多发电机决策产生的高维决策空间，而增量式由于是在已有合适决策基础上修正，训练难度相较于覆盖式低一些。因而，可以预计智能体采用增量式 BDQ 结构具有最好的训练效果。

4）智能体训练参数设置：智能体的神经网络训练总轮次 N_T=30 000，经验池的容量为 D=10 000，从经验池抽取批样本的容量为 B=128，目标 Q 网络的参数同步更新周期 K=20，折扣系数γ=0.1，学习速率α=10^{-4}，贪婪系数的初始值为ε_s=0.9，最低值为ε_e=0.05，衰减系数 n_d=4 000，每个回合最大迭代次数为 N_L=20，切机量的惩罚系数为λ_1=−0.1，切机决策有效的奖励λ_2=10。智能体的神经网络使用 PyTorch 框架编写并启用 CUDA 加速训练，智能体通过 Python 调用 PSASP 仿真软件执行切机决策。

4.2 智能紧急切机的训练和决策效果

1）智能体切机决策实例：以一个实际测试工况来验证智能体的切机决策效果。如图 10(a)所示，在预想故障工况的测试集合中选择一个工况，其潮流水平为[1.1, 1.1, 1.15, 1.15]，线路 23-24 靠近节点 23 的 10%处发生三相短路接地故障，故障持续 0.25 s。在没有施加切机决策之前，系统功角失稳，智能体根据功角和电压信息输入给出决策：节点 G5 切 4 台、节点 G6 切 4 台、节点 G7 切 4 台。如图 10(b)所示，施加切机决策后最终系统恢复功角稳定，由于 BDQ 网络具有多节点同时决策的能力，所以可以 1 次给出有效决策，减少了切机的尝试次数和耗时。

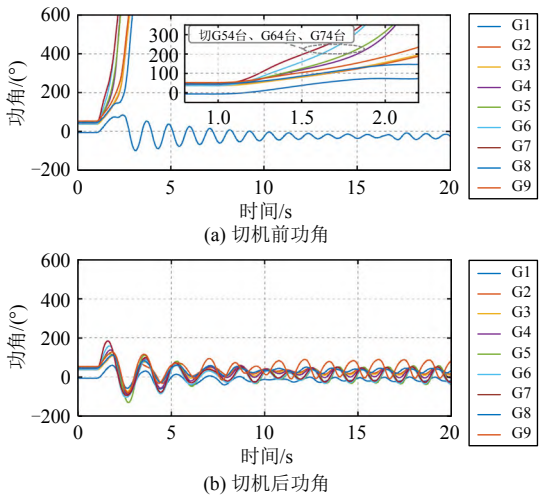


图 10 训练好后的智能体执行紧急切机决策制定

Fig. 10 Trained agent performs emergency generator rejection decision-making

2）智能体的决策性能指标：为了评估训练完成后的智能体决策性能，用总迭代次数、总决策成功次数和总回报 3 个性能指标进行评估，指标含义和与决策性能之间的关系参考表 2。

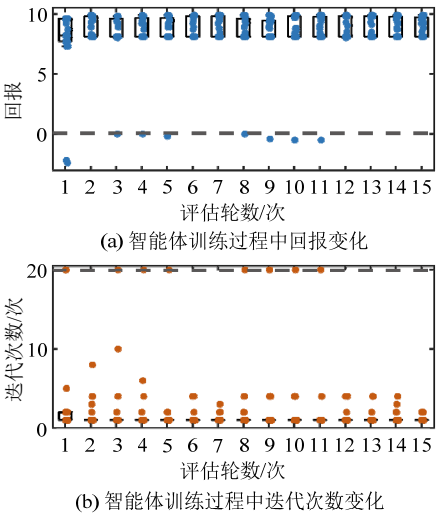
表 2　评估智能体决策性能的指标

Table 2　Evaluation index of agent decision performance

指标	含义说明	与决策性能的关系
总迭代次数	智能体在所有工况下决策尝试次数	越小越好
总决策成功次数	智能体给出有效决策的工况数量	越大越好
总回报	智能体在所有工况下获得的回报之和	越大越好

总迭代次数对应于总仿真次数，因此该指标反映了智能体决策的快速性。总决策成功次数反映了智能体给出有效决策的能力。总回报的大小与切机决策量的大小相关，在决策有效的前提下决策量越小，获得的回报越大，因此该指标反映了智能体决策质量的高低。

3）智能体训练过程中的性能变化：如图 11 所示，在 30 000 个训练轮次中，每隔 2 000 次对智能体的决策性能进行评估，一共进行 15 次评估。每一轮对所有的测试集合上的预想故障工况进行评估得到对应的一列点，每一个点表示一个工况，点的分布表征了智能体经过训练后的决策性能，每一个黑色方框上下边是箱形图标记的上下 1/4 分位数。下面以增量式 BDQ 智能体为例，评估智能体在测试集合上每个故障工况下的回报和迭代次数。理想情况下的回报极限是 10，即没有切机决策造成的惩罚，并获得切机决策成功得到的奖励。如果回报低于 0 则说明智能体在这个测试工况下没有能给出有效决策。在训练过程中，随着训练的进行，智能体在决策制定过程中所得的回报平均值上升，迭代次数逐渐下降，最后每个测试工况都能给出有效决策，且决策的迭代次数都在 2、3 次左右。



(a) 智能体训练过程中回报变化

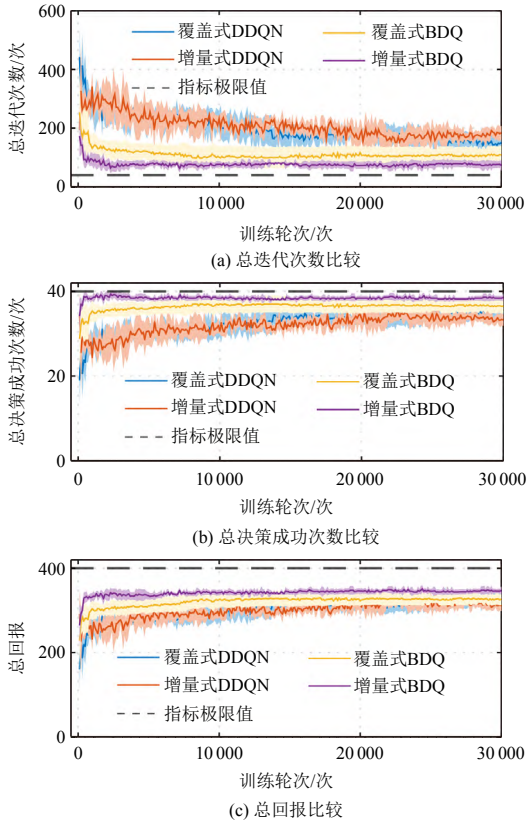
(b) 智能体训练过程中迭代次数变化

图 11　智能体在测试集合上的决策性能

Fig. 11　Decision performance of agent on test set

4.3　不同动作和决策空间结构的对比

3 节介绍了 4 种 DQN 结构，为了比较这几种结构的优劣，对智能体在测试故障工况集合上进行性能评估，每 2 000 个训练轮次就进行一次评估，每种结构都随机重复执行 5 次，结果如图 12 所示。实线表示多次结果的平均值，浅色块表现性能指标的方差，黑色虚线表示在完全理想状态下各最优性能的指标极限值，此时，对于任意工况只需要 1 次尝试即可得到最优有效决策，总迭代次数最低 40 次，总成功次数最高 40 次，总回报的上界是 400。



(a) 总迭代次数比较

(b) 总决策成功次数比较

(c) 总回报比较

图 12　4 种 DQN 算法的决策性能比较

Fig. 12　Comparison of decision-making performance of four DQN algorithms

如图 12 所示，对比 BDQ 算法和 DDQN 算法，不论是覆盖式还是增量式，使用 BDQ 结构能够显著提升智能体的训练收敛速度和训练收敛后的决策性能。从原理上理解，使用 BDQ 的多分支特性使得智能体在同时决策的探索中更容易先找到有效决策，然后在训练的过程中渐渐提升决策智能体的性能，因此在决策智能体的构建中使用 BDQ 结构比 DDQN 结构更好。

然后对比使用覆盖式和增量式两种动作-决策转换关系的性能差距，可以发现对于 DDQN 结构来

说差别不大，而对于 BDQ 结构来说，使用增量式结构比覆盖式结构能够进一步提升智能体的性能。从原理上理解，使用增量式结构只需要在已有决策上进行一定的修正，通常更容易找到有效决策，所以相对降低了算法估计决策量的难度。综合来看，在本文所提智能紧急切机决策制定算法中智能体使用增量式 BDQ 结构能够以更快的收敛速度、更好的决策有效性和决策质量实现切机决策制定。

4.4 知识经验引导探索的验证

知识经验引导在智能体训练过程中起作用，而且无需参数设置，算例设置与前面一致。如图 13(a)所示，智能体决策总迭代次数在使用知识经验引导探索后可以从 72 左右下降至 52 左右，比较接近于理想最低决策迭代次数 40，大部分测试集上决策成功的失稳工况都是一次给出有效的切机决策。

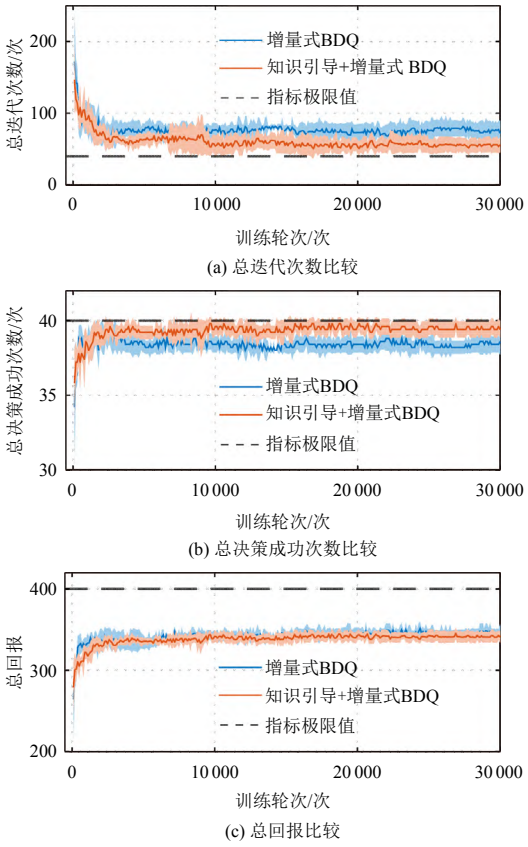


图 13 采用知识经验引导前后决策性能对比

Fig. 13 Comparison of decision-making performance whether to use knowledge and experience guidance

从图 13(b)可以发现，总决策成功次数也基本稳定在 39~40 次，相比于没有使用知识经验引导训练的智能体决策性能更好。从图 13(c)可以发现，使用知识经验引导策略前后的总回报差距不大。因此，使用知识经验引导有利于提升增量式 BDQ 智能体训练收敛的总迭代次数和决策成功次数。

4.5 多样本源混合训练的验证

为验证使用多样本源混合采样训练策略后提升智能体决策性能的有效性，按照式(27)设置起始优质样本占比 c_s 为 0.9，优质样本指数衰减参数 n_c 为 4 000，最终优质样本占比 c_e 为 0.3。如图 14 所示，在采用多样本源混合采样训练策略后，在总迭代次数和决策成功次数上略有优势，总回报从原来的 338 左右提升至 369 左右。

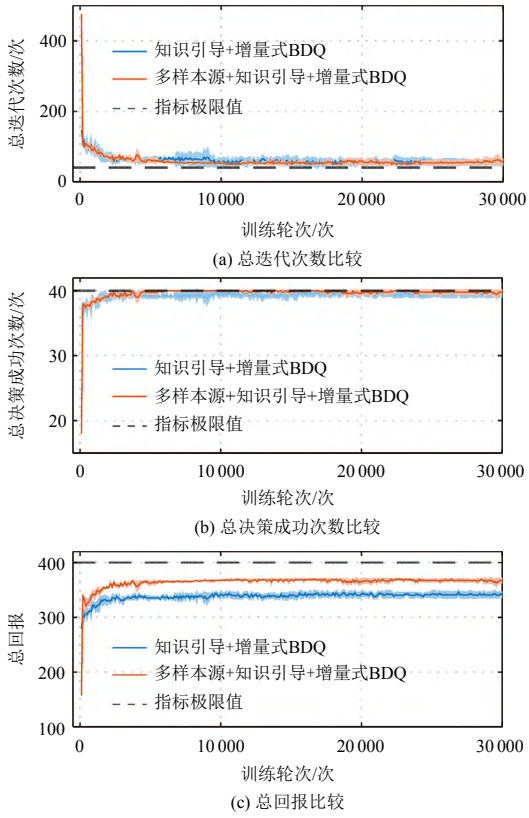


图 14 采用多样本源混合采样训练前后决策性能对比

Fig. 14 Comparison of decision performance on whether to use multi-source mixed sampling training

从原理上理解，引入优质样本后，智能体在训练过程中充分学习优质样本中决策的特点，可以提升决策质量，即以较小的切机成本使系统恢复功角稳定。但就迭代次数和决策成功次数来说，智能体的性能已接近于极限值，可提升空间较小。

4.6 不同策略的泛化能力和鲁棒性的验证

为了验证不同策略的泛化性能以及鲁棒性，除了前述训练集和测试集划分外，又将整个数据集随机做 5 次划分，每次划分做 5 次重复仿真计算各项性能取平均值，然后再对 6 次划分求得平均值和标准差。各种策略的性能统计整理如表 3 所示，从表 3 可以看出，各种算法的性能指标方差相对较小，

表 3　不同训练集划分下不同训练策略的性能统计

Table 3　Performance statistics of different training strategies under different training set divisions

测试策略	总回报	总成功次数	总迭代次数
增量式 BDQ	354.12±6.50	38.63±0.43	73.09±11.93
知识引导+增量式 BDQ	350.63±9.16	39.52±0.38	53.96±8.70
多样本源+知识引导+增量式 BDQ	373.04±3.86	39.83±0.15	51.01±6.14

说明各算法具有较强的泛化性能。另外，使用两种知识融合策略后决策性能提升，方差逐步下降，说明知识融合有助于提升算法的鲁棒性。

4.7　智能体的训练及决策时间成本

本文所提算法的实际训练时间平均在 20~30 min，如果保存已经仿真过的样本数据供后续快速读取，则训练时间可以进一步缩短。决策时间由决策时间和决策步数共同决定，从表 3 可知，在 40 个测试工况下，总迭代次数即总仿真次数平均约为 51 次，平均 1 个工况约需要 1.3 次仿真。而单次仿真时间的长短与仿真系统的规模以及计算机算力、仿真软件的求解算法效率等相关。但本文的主要目标是在不考虑单次仿真时长的情况下，尽可能降低总迭代次数来实现紧急切机决策的快速制定。综合来看，本文所提算法能有效降低训练及决策时间成本，以实现紧急切机决策的快速制定。

5　结论

本文提出一种基于知识融合和深度强化学习的智能紧急切机决策制定框架，针对紧急切机问题设计马尔可夫决策过程以及智能体的动作-决策空间。智能体通过交互产生样本并不断训练更新网络，训练完成后可对给定工况直接输出有效的切机决策。

在智能体结构设计方面，当需要处理多个可控发电机同时决策产生的高维决策空间时，可以采用决策空间压缩或者应用 BDQ 网络。另外，智能体的动作-决策转换关系包括覆盖式和增量式两种。算例结果表明，使用增量式 BDQ 结构相比其他结构在决策总迭代次数、总成功次数、总回报 3 个指标上具有更优性能。

为进一步提升智能体性能，本文提出两种知识融合策略。算例结果表明，用知识经验引导策略可以提升决策成功次数，减少决策迭代次数；用多样本源混合训练策略可以让智能体学习已有的优质样本的特点提升决策质量。

下一步会在所提智能紧急切机决策框架基础上改进，融合更多特定运行经验缩减决策空间，依托更强算力在大规模的实际电网中进行验证。

参考文献

[1] 汤涌．电力系统安全稳定综合防御体系框架[J]．电网技术，2012，36(8)：1-5．
TANG Yong．Framework of comprehensive defense architecture for power system security and stability [J]. Power System Technology，2012，36(8)：1-5 (in Chinese).

[2] 胡伟，张玮灵，闵勇，等．基于支持向量机的电力系统紧急控制实时决策方法[J]．中国电机工程学报，2017，37(16)：4567-4576．
HU Wei，ZHANG Weiling，MIN Yong，et al．Real- time emergency control decision in power system based on support vector machines[J]．Proceedings of the CSEE，2017，37(16)：4567-4576 (in Chinese).

[3] 王怀远，张保会，杨松浩，等．电力系统暂态稳定切机控制策略表的快速整定方法[J]. 电力系统自动化，2016，40(11)：68-72，79．
WANG Huaiyuan，ZHANG Baohui，YANG Songhao，et al．Fast setting method of generator tripping strategy tables in transient stability control of power systems [J]. Automation of Electric Power Systems，2016，40(11)：68-72，79 (in Chinese).

[4] YANG Songhao，HAO Zhiguo，ZHANG Baohui，et al．An accurate and fast start-up scheme for power system real-time emergency control[J]．IEEE Transactions on Power Systems，2019，34(5)：3562-3572．

[5] 滕林，刘万顺，贠志皓，等．电力系统暂态稳定实时紧急控制的研究[J]．中国电机工程学报，2003，23(1)：64-69．
TENG Lin，LIU Wanshun，YUN Zhihao，et al．Study of real-time power system transient stability emergency control[J]．Proceedings of the CSEE，2003，23(1)：64-69 (in Chinese).

[6] 吴为，汤涌，孙华东．基于系统加速能量的切机控制措施量化研究[J]．中国电机工程学报，2014，34(34)：6134-6140．
WU Wei，TANG Yong，SUN Huadong．Quantitative research of generation capacity tripped based on acceleration energy of power system[J]．Proceedings of the CSEE，2014，34(34)：6134-6140 (in Chinese).

[7] 李志浩．大规模电力系统暂态稳定紧急控制研究[D]．杭

州：浙江大学，2017.
LI Zhihao. Research on emergency control of transient stability in large power systems[D]. Hangzhou：Zhejiang University，2017(in Chinese).

[8] TANG Yong，HUANG Yanhao，WANG Hongzhi，et al. Framework for artificial intelligence analysis in large-scale power grids based on digital simulation [J]. CSEE Journal of Power and Energy Systems，2018，4(4)：459-468.

[9] 侯玉强，崔晓丹，李威，等. 用于在线系统的安控策略优化搜索方法[J]. 中国电机工程学报，2011，31(S1)：73-76.
HOU Yuqiang，CUI Xiaodan，LI Wei，et al. A new method of searching optimal stability control strategy for on-line system[J]. Proceedings of the CSEE，2011，31(S1)：73-76 (in Chinese).

[10] 张保会，王怀远，杨松浩，等. 电力系统暂态稳定性闭环控制(五)——控制量的实时计算[J]. 电力自动化设备，2014，34(12)：1-5.
ZHANG Baohui，WANG Huaiyuan，YANG Songhao，et al. Closed-loop control of power system transient stability(5)：calculation of control quantity[J]. Electric Power Automation Equipment，2014，34(12)：1-5 (in Chinese).

[11] 任伟，房大中，陈家荣，等. 大电网暂态稳定紧急控制下切机量快速估计算法[J]. 电网技术，2008，32(19)：10-15，55.
REN Wei，FANG Dazhong，CHEN Jiarong，et al. A fast algorithm to estimate generation capacity tripped by emergency control for transient stability of large power system[J]. Power System Technology，2008，32(19)：10-15，55 (in Chinese).

[12] 王彪，方万良，罗煦之. 紧急控制下最优切机切负荷方案的快速算法[J]. 电网技术，2011，35(6)：82-87.
WANG Biao，FANG Wanliang，LUO Xuzhi. A fast algorithm of optimal generator and load-shedding for emergency control[J]. Power System Technology，2011，35(6)：82-87 (in Chinese).

[13] 陆崎，任祖怡，徐柯，等. 基于隐枚举法的稳定控制优化切机方法[J]. 电力系统自动化，2016，40(5)：139-144.
LU Qi，REN Zuyi，XU Ke，et al. Optimal generator tripping scheme based on implicit enumeration method[J]. Automation of Electric Power Systems，2016，40(5)：139-144 (in Chinese).

[14] 毕兆东. 电力系统暂态稳定控制决策算法[D]. 杭州：浙江大学，2002.
BI Zhaodong. Algorithm of transient stability emergency control in power system[D]. Hangzhou：Zhejiang University，2002 (in Chinese).

[15] MNIH V，KAVUKCUOGLU K，SILVER D，et al. Human-level control through deep reinforcement learning[J]. Nature，2015，518(7540)：529-533.

[16] LILLICRAP T P，HUNT J J，PRITZEL A，et al. Continuous control with deep reinforcement learning[C]//4th International Conference on Learning Representations. San Juan：ICLR，2016.

[17] YAN Ziming，XU Yan. Data-driven load frequency control for stochastic power systems：a deep reinforcement learning method with continuous action search[J]. IEEE Transactions on Power Systems，2019，34(2)：1653-1656.

[18] 刘威，张东霞，王新迎，等. 基于深度强化学习的电网紧急控制策略研究[J]. 中国电机工程学报，2018，38(1)：109-119.
LIU Wei，ZHANG Dongxia，WANG Xinying，et al. A decision making strategy for generating unit tripping under emergency circumstances based on deep reinforcement learning[J]. Proceedings of the CSEE，2018，38(1)：109-119 (in Chinese).

[19] HUANG Qiuhua，HUANG Renke，HAO Weituo，et al. Adaptive power system emergency control using deep reinforcement learning[J]. IEEE Transactions on Smart Grid，2020，11(2)：1171-1182.

[20] HUANG Renke，CHEN Yujiao，YIN Tianzhixi，et al. Learning and fast adaptation for grid emergency control via deep meta reinforcement learning[J]. arXiv：2101.05317，2022.

[21] 严梓铭，徐岩. 结合深度强化学习与领域知识的电力系统拓扑结构优化[J]. 电力系统自动化，2022，46(1)：60-68.
YAN Ziming，XU Yan. Topology optimization of power systems combining deep reinforcement learning and domain knowledge[J]. Automation of Electric Power Systems，2022，46(1)：60-68 (in Chinese).

[22] 国家市场监督管理总局，国家标准化管理委员会. GB 38755—2019 电力系统安全稳定导则[S]. 北京：中国标准出版社，2019.
State Administration for Market Regulation，Standardization Administration of the People's Republic of China. GB 38755—2019 Code on security and stability for power system[S]. Beijing：Standards Press of China，

2019 (in Chinese).

[23] SHI Zhongtuo，YAO Wei，ZENG Lingkang，et al．Convolutional neural network-based power system transient stability assessment and instability mode prediction[J]．Applied Energy，2020，263：114586．

[24] TAVAKOLI A，PARDO F，KORMUSHEV P．Action branching architectures for deep reinforcement learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence．New Orleans，USA：AAAI，2018．

[25] VAN HASSELT H，GUEZ A，SILVER D．Deep reinforcement learning with double q-learning[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence．Phoenix：ACM，2016：2094-2100．

[26] WANG Ziyu，SCHAUL T，HESSEL M，et al．Dueling network architectures for deep reinforcement learning[C]//Proceedings of the 33rd International Conference on Machine Learning．New York，NY，USA：ACM，2016：1995-2003．

[27] 张保会，王怀远，杨松浩．电力系统暂态稳定性闭环控制(六)——控制地点的选择[J]．电力自动化设备，2015，35(1)：1-5，12．
ZHANG Baohui，WANG Huaiyuan，YANG Songhao．Closed-loop control of power system transient stability (6)：control location selection[J]．Electric Power Automation Equipment，2015，35(1)：1-5，12 (in Chinese)．

李舟平

在线出版日期：2023-01-11。
收稿日期：2022-09-27。

作者简介：
李舟平(1997)，男，工学硕士，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，lizp_hust@163.com；
*通信作者：姚伟(1983)，男，工学博士，教授，研究方向为新能源电力系统的稳定性分析与控制，新一代电力人工智能技术及应用，w.yao@hust.edu.cn。

(责任编辑　邱丽萍)

 DOI: 10.13334/j.0258-8013.pcsee.222633

Intelligent Emergency Generator Rejection Schemes Based on Knowledge Fusion and Deep Reinforcement Learning

LI Zhouping[1], ZENG Lingkang[1], YAO Wei[1*], HU Ze[1], SHUAI Hang[2], TANG Yong[3], WEN Jinyu[1]

(1. State Key Laboratory of Advanced Electromagnetic Engineering and Technology (School of Electrical and Electronic Engineering, Huazhong University of Science and Technology); 2. Department of Electrical Engineering and Computer Science, University of Tennessee; 3. China Electric Power Research Institute)

KEY WORDS: emergency generator rejection decision; deep reinforcement learning; decision space; branching dueling Q network; knowledge fusion

Emergency control is an important means of maintaining power system transient security and stability following serious faults. The current popular "human-in-the-loop" offline emergency control decision-making method has some drawbacks, including low efficiency and heavy reliance on expert experience.

This paper proposes an intelligent emergency generator rejection decision-making method based on knowledge fusion and deep reinforcement learning (DRL). Fig. 1 depicts the framework of it. Through interactions with the power system environment, the agent updates its parameters using generated samples to improve its decision ability gradually.

When the typical DRL agent deals with multi-generator decisions, the resulting high-dimensional decision space makes the agent's training difficult. There are two solutions proposed: decision space compression and the use of a branching dueling Q (BDQ) network.

To further improve the exploration efficiency and decision quality of the agent, the knowledge and experience related to emergency generator rejection control are integrated to the agent training. Specifically, two knowledge fusion strategies, knowledge guidance and multi-source training, are adopted in this paper.

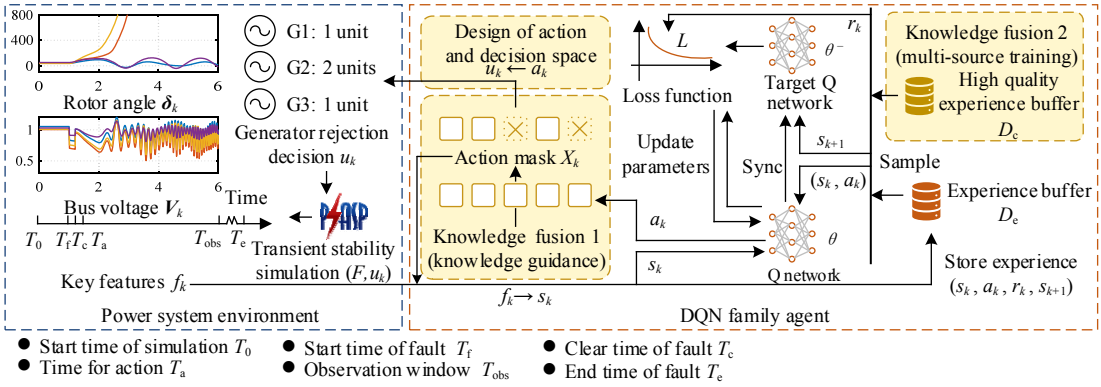


Fig. 1 Framework of intelligent emergency generators rejection decision for power system based on DRL with knowledge fusion

The simulation results of 10-machine 39-bus system show that the fully trained agent can quickly give effective emergency generator rejection decisions.

In dealing with the high dimensional decision space problem, applying both a BDQ network and decision space compression are effective. In addition, the action-decision transformation of the agent includes two types: overwritten type and incremental type. The simulation results show that the incremental BDQ agent has better performance than other agents.

Moreover, several tests have been conducted to verify the effectiveness of the knowledge fusion strategy. The results are shown in Table 1. Since the knowledge guidance strategy reduces ineffective explorations, the mean of total decision iterations is reduced from about 73 to 53, which is close to the minimum of 40. The multi-source training strategy helps the agent learn useful characteristics from high-quality samples, and the mean of total return is increased from about 354 to 373. Hence, the knowledge fusion strategy does improve the exploration efficiency and decision quality.

Table 1 Performance statistics of different training strategies under different training set divisions

Testing strategies	Total return	Total successes	Total decision iterations
incremental BDQ	354.12 ± 6.50	38.63 ± 0.43	73.09 ± 11.93
incremental BDQ with knowledge guidance	350.63 ± 9.16	39.52 ± 0.38	53.96 ± 8.70
incremental BDQ with knowledge guidance and multi-source training	373.04 ± 3.86	39.83 ± 0.15	51.01 ± 6.14