

结合深度强化学习与领域知识的电力系统拓扑结构优化

严梓铭, 徐 岩

(南洋理工大学电气与电子工程学院, 新加坡 639798, 新加坡)

摘要: 对拓扑结构进行优化可提高电力系统运行灵活性,然而线路开断与变电站母线分裂等系统级的离散决策变量维度极高。该拓扑结构优化问题难以由传统混合整数优化方法求解。针对该问题,提出了一种结合异步优势 Actor-Critic(A3C)深度强化学习与电力系统领域知识的运行优化方法,将在线优化的计算负担转移至离线智能体训练阶段。该方法通过同时考虑拓扑结构与发电出力调整的动作空间设计系统运行控制智能体,以最小化约束越限为训练奖励,通过强制约束校验缩减搜索空间并提高强化学习效率,从而实现电力系统运行拓扑结构优化的快速计算,提高电力系统运行的安全性。仿真测试验证了所提方法的有效性。

关键词: 电力系统运行; 深度强化学习; 拓扑结构优化; 发电出力调整

0 引言

电力系统网络结构一般不随系统运行状态改变而改变,若非检修或故障,线路等主要输电设备将保持闭合运行^[1],而电力系统运行调度也通常不考虑拓扑结构调整。近年来,随着电力系统不确定性日趋复杂,可再生能源并网比例提高,传统电力系统发电出力调整的灵活性有限,难以在有限控制资源下满足电力系统安全经济运行的要求。因此,有研究开始尝试将电网拓扑结构作为电力系统运行的控制变量^[2],以进一步优化电力系统的运行水平。

输电网拓扑结构优化无须加装设备,可通过改变电网拓扑来改善潮流分布,从而解决线路过载与电压越限等问题。近年来,常见的输电网结构优化主要包括输电线路最优开断^[3-12]和变电站母线分裂^[13-17]这2类。在线路最优开断方面,文献[3]提出了基于鲁棒优化与故障遍历校验的网络结构优化方法。文献[4]计及电网 $N-1$ 安全性与短路电流,采用混合整数线性规划求解线路开断问题。文献[5]提出了基于灵敏度分析、排序与验算对比的网络拓扑调整方法。文献[6]提出了改善暂态稳定性的线路开断方法。文献[7]基于直流潮流模型求解最优传输线开断问题。文献[8-10]采用启发式算法求解含传输线开断的最优潮流模型。文献[11]采用混合整数非线性规划求解最优传输线开断问题。在变电

站母线分裂方面,文献[13]介绍了变电站母线分裂等拓扑结构控制的概念,并采用混合整数规划求解考虑变电站拓扑的安全约束最优潮流模型。文献[14]提出了一种考虑最优传输线与变电站母线的电网拓扑结构优化方法。文献[15]将母线分裂优化问题建模为混合整数优化问题,并实现了个位数变电站的母线分裂求解。文献[16]采用输电网结构混合整数非线性规划求解得到改善电网均匀度的方案。上述文献中主要通过较小的变量集上计算混合整数优化问题来求解拓扑结构,然而,在实际变电站中,各设备在母线上的不同连接方式都可作为独立的拓扑决策,考虑所有变电站的系统级拓扑决策变量繁多,难以由传统混合整数优化方法在线求解该超高维度拓扑结构优化问题。此外,混合整数优化方法仅改善当前时刻的系统运行成本,忽略了未来系统运行的状态。相对而言,深度强化学习可通过价值函数近似考虑未来电力系统的运行状态^[18],可在考虑线路开断^[19]与不同拓扑结构^[20-21]等离散变量的情况下快速求解最优控制问题,有望前瞻性地降低安全隐患,但存在图论模型灵活性低或搜索空间和学习负担过大的问题。

为实现系统级拓扑结构优化问题的求解并降低深度强化学习负担,本文提出了一种结合异步优势 Actor-Critic(A3C)深度强化学习与领域知识的电力系统拓扑结构优化方法。以提升系统在各种随机场景下的 $N-1$ 安全性为核心,本文采用最小化约束越限为奖励训练智能体,将在线运行的优化计算负

收稿日期: 2021-05-10; 修回日期: 2021-09-04。

上网日期: 2021-12-02。

担转移至离线训练过程,同时通过动作空间筛选降低训练负担,实现了良好的学习效果。

1 考虑拓扑变量的电力系统运行优化模型

1.1 问题的描述与建模

对于实际电网,每个变电站均有多条母线。每一个负荷、每一台发电机或每一条传输线均可连接在其中一条母线或另一条母线上。通过变电站节点分裂或重连、设备与线路所连母线的变化、线路开断,都可以改变电力系统的拓扑结构,从而改变潮流分布,实现系统的经济与安全运行。

在考虑拓扑的电力系统运行优化问题中,要求控制中心尽可能让系统在负荷波动、线路故障等随机因素影响下安全运行更长的时间,并降低运行成本。过载与线路故障可能造成连锁故障并使潮流发散,因此系统拓扑结构优化的问题本质是考虑 $N-1$ 安全性的最优潮流问题。

计及传输线网损和发电出力调整经济补偿,并尽可能避免停电,系统运行的总成本可表述为:

$$C \approx \sum_{t=1}^{T_g} (c_l(t) + c_r(t)) + \sum_{t=T_g}^{T_e} c_b(t) \quad (1)$$

$$c_l(t) \approx p(t) \sum_{l=1}^{n_l} r_l y_l^2(t) \quad (2)$$

$$c_r(t) \approx 2\alpha p(t) \varepsilon_r(t) \quad (3)$$

$$c_b(t) \approx D(t) \beta p(t) \quad (4)$$

式中: C 为单个场景的总运行成本; $c_l(t)$ 为时刻 t 传输线网损成本; $p(t)$ 为时刻 t 电价; r_l 为传输线 l 的电阻; $y_l(t)$ 为传输线 l 在时刻 t 的电流; $c_r(t)$ 为时刻 t 发电机出力调整的成本; $c_b(t)$ 为时刻 t 停电成本; α 为补偿系数,因增发或少发电的发电厂都有偏离计划的出力调整,因此电网运营商需按补偿协议 $\alpha \geq 1$ 弥补双方的损失; $\varepsilon_r(t)$ 为发电出力调整量; $D(t)$ 为时刻 t 总负荷; β 为停电成本系数, $\beta \geq 1$; T_g 为系统正常运行总时长; T_e 为系统停电总时长; n_l 为传输线总数。

为实现总成本最小化,系统须满足潮流方程式约束(5)和(6),以及发电机有功出力约束(7)、机组爬坡速率约束(8)、线路潮流约束(9)、线路热稳定极限约束(10)等不等式约束。

$$P_{G,i} - P_{D,i} = V_i \sum_{j=1}^n V_j (G_{ij} \cos \delta_{ij} + B_{ij} \sin \delta_{ij}) \quad (5)$$

$$Q_{G,i} - Q_{D,i} = V_i \sum_{j=1}^n V_j (G_{ij} \sin \delta_{ij} - B_{ij} \cos \delta_{ij}) \quad (6)$$

$$P_{G,i}^{\min} \leq P_{G,i}^t \leq P_{G,i}^{\max} \quad (7)$$

$$P_{G,i}^{t-1} - R_{G,i}^{\text{down}} \leq P_{G,i}^t \leq P_{G,i}^{t-1} + R_{G,i}^{\text{up}} \quad (8)$$

$$|S_{ij}| \leq S_{ij}^{\max} \quad (9)$$

$$|S_{ij}| < S_{ij}^{T,\max} \quad (10)$$

式中: $P_{D,i}$ 、 $Q_{D,i}$ 、 $P_{G,i}$ 、 $Q_{G,i}$ 分别为节点 i 的有功负荷、无功负荷、有功出力、无功出力; V_i 和 V_j 分别为节点 i 和 j 的电压幅值; δ_{ij} 为母线 i 与母线 j 之间的相角差; $Y_{ij} = G_{ij} + jB_{ij}$ 为节点导纳矩阵第 i 行 j 列元素, G_{ij} 和 B_{ij} 分别为相应元素的电导和电纳; $P_{G,i}^t$ 为时刻 t 发电机 i 的有功出力; $R_{G,i}^{\text{down}}$ 和 $R_{G,i}^{\text{up}}$ 分别为机组向下和向上爬坡率; S_{ij} 为输电线路(i,j)的传输功率; S_{ij}^{\max} 为输电线路(i,j)的最大额定值功率; $S_{ij}^{T,\max}$ 为输电线路(i,j)热稳定极限允许的最大功率; n 为系统总节点数。

在系统运行时,由于受运行约束与热稳定极限约束,当线路过载超过一定时间后就会断开,而当线路潮流超过热稳定极限约束时就会立即断开。在智能体动作后,通过求解潮流方程判断潮流方程是否发散,若发散则判断为开始停电,从而可统计总运行成本。

1.2 基于深度强化学习的系统运行优化

拓扑结构的决策是一个含离散变量的高维优化问题,而每种不同的拓扑决策不仅影响当前状态下的潮流分布,还会影响若干时间后系统在其他状况下的潮流分布。因此,传统优化问题难以在考虑未来系统状态的情况下及时求解最优决策问题。为此,可将考虑拓扑变量的电力系统运行优化问题建模成马尔可夫决策过程,并通过深度强化学习来离线训练智能体,在线应用时能及时求解系统最优运行状态。

目前,主流的深度强化学习方法包括基于价值(value-based)和基于策略(policy-based)的深度强化学习方法,前者会评价不同动作的预期收益并选择最优收益所对应的动作,而后者直接求得使目标最优的动作概率(或值)。两种深度强化学习均通过与环境迭代互动来优化智能体。基于深度强化学习优化系统运行的过程可以用图1表示。如图1所示,智能体将观测电力系统运行状态,并求解最优控制动作;环境通过仿真对智能体动作进行评价,从而反馈给智能体,以引导智能体沿最大化预期奖励的方向进行强化学习。

2 结合领域知识并基于A3C的电力系统运行深度强化学习策略

2.1 A3C深度强化学习

由于电力系统拓扑决策的动作空间巨大,单线

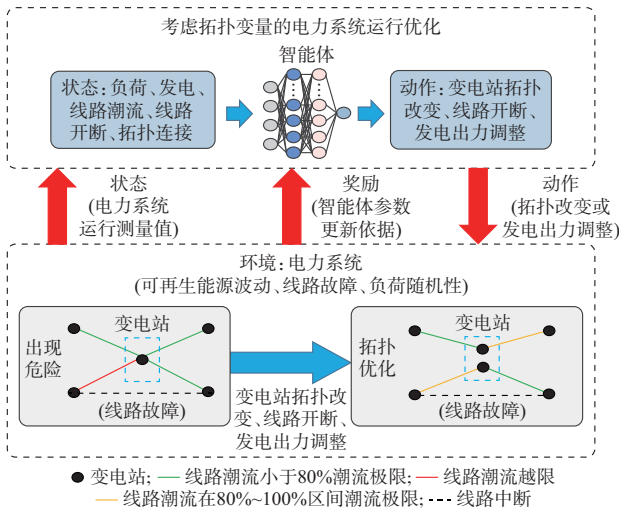


图1 用于电力系统拓扑优化的深度强化学习过程
Fig.1 Deep reinforcement learning process for power system topology optimization

程的强化学习策略难以有效遍历可行决策。为了增加样本多样性,可设置多个线程,令智能体分别在学习过程中与不同的环境交互,从而克服难以收敛的问题。为此,本文基于A3C算法设计电力系统运行控制智能体,通过并行创建多个不同的环境,让多个不同的智能体同时并在各个环境中更新全局网络的参数,从而增加样本的多样性并改善强化学习的收敛性。训练过程框架如图2所示,通过设置不同的环境场景,分别在不同线程中运行智能体,A3C网络结构使用系统状态作为输入(详见2.2节),在经过共享隐含层后,经由Actor隐含层计算并输出动作空间中各个动作(详见2.2节)的概率,Actor会以概率采样或直接选择最高概率动作并执行。共享隐含层之后,Critic会评价在当前状态下的状态价值。

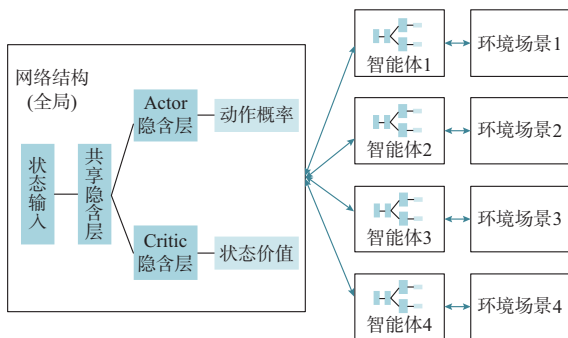


图2 用于电力系统运行优化的A3C强化学习智能体训练框架
Fig.2 Training framework of A3C reinforcement learning agent for power system operation optimization

在以上异步框架的基础上,A3C通过优势项,即智能体的实际奖励与Critic输出预期状态价值之

间的差异来加快与稳定训练过程。其训练过程主要由式(11)和式(12)更新深度神经网络参数:

$$D\theta = D\theta + \nabla_{\theta'} \log(\pi(a_i|s_i; \theta'))(R - V(s_i; \theta'_v)) \quad (11)$$

$$D\theta'_v = D\theta'_v + \frac{\partial((R - V(s_i; \theta'_v))^2)}{\partial \theta'_v} \quad (12)$$

式中: a_i 为智能体*i*的控制动作(即拓扑决策与发电出力调整); s_i 为智能体*i*的状态(即电力系统当前拓扑结构与潮流分布); $V(s_i; \theta'_v)$ 为预期状态价值,即Critic神经网络预测的状态价值; R 为奖励函数; θ 为Actor神经网络参数,由多个运行过程下的优势项对Actor参数梯度的累计求和更新; θ'_v 为Critic神经网络参数,由状态价值预测误差对Critic参数梯度的累计求和更新; π 为智能体动作策略(policy); θ' 和 θ'_v 分别为目标Actor和Critic神经网络参数。

经过离线的深度强化学习,智能体参数 θ 可以得到充分更新,并最大化运行过程中的预期奖励。在此基础上,以电力系统运行状态为输入、拓扑决策或发电出力调整的动作概率为输出,所得的A3C深度强化学习智能体可用于在线的电力系统运行优化。在训练完成后,智能体可实时根据电力系统状态计算得到当前时刻各动作的概率,概率最高的动作将会作为智能体的决策。在智能体选择动作后,本文由潮流计算进一步检验动作的可行性。若当前动作无法满足所有约束条件,则继续检验概率次高的拓扑改变或发电出力调整动作,直至动作满足约束条件为止。

2.2 特征与动作空间筛选

在将电力系统运行优化问题建模为深度强化学习问题的过程中,特征(智能体的输入)与动作(智能体的可选输出)的选择直接决定了性能的好坏。

2.2.1 特征筛选

本文在智能体建模过程中使用的特征包括数值特征与拓扑结构特征。数值特征包括负荷有功功率和无功功率、发电有功功率和无功功率、预测负荷、线路维修计划(距下次维修所剩小时数)、线路功率与功率约束。拓扑结构特征基于节点临界矩阵,在生成无向图的基础上,由NetworkX包计算得到,包括图的节点重要性(PageRank)、介数中心性(betweenness centrality)、节点度数(degree centrality)。线路开断状态直接作为额外的向量并入拓扑特征中。需要指出的是,图神经网络^[20]理论上更适合用于拓扑结构优化问题。在此基础上,数值特征与拓扑结构特征可同时存入列向量中,作为全连接神经网络(智能体)的输入。

2.2.2 动作筛选

原问题可控参数包括系统拓扑结构与发电机再调度功率,动作空间维度过于巨大,难以直接训练强化学习智能体。受限于机组爬坡速率约束,本文将所有的机组再调度问题都建模为离散决策问题。设机组再调度增发或减发功率为机组爬坡率所限制功率,发电再调度模型可建模为从 n_g 台可调整发电机中选择 k 台发电机进行增发或减发功率(共 $C_{2n_g}^k$ 种不同决策),可将此离散决策动作采用不放回采样的排列组合问题求解得到。通过排列组合工具箱Itertools可直接枚举所有包含发电机的决策组合,并在此基础上通过动作筛选缩减动作空间。

在枚举所有可行拓扑决策与再调度决策的基础上,本文使用了滚动式动作筛选。设置一个基于穷举法的智能体在环境中不断运行,仅当系统发生潮流越限时开始筛选可行动作,并不断仿真直到迭代结束。在动作筛选过程中,能让系统从潮流越限状态恢复至安全运行状态的拓扑动作与发电出力调整动作合并至智能体训练所用的动作空间:

$$\Omega_{A,h+1} = \Omega_{A,h} \cup (\Omega_{S,h} \cap \Omega_{T,h}) \cup (\Omega_{S,h} \cap \Omega_{R,h}) \quad (13)$$

$$\begin{cases} F_h(a_i) < F_{h,\max} \\ R(a_i) > R(a_0) \end{cases} \quad \forall a_i \in \Omega_{S,h} \quad (14)$$

$$\Omega_{R,h} = \Omega_R(C_{2n_g}^1) \cup \Omega_R(C_{2n_g}^2) \cup \Omega_R(C_{2n_g}^3) \quad (15)$$

式中: $\Omega_{A,h}$ 和 $\Omega_{A,h+1}$ 分别为第 h 步和 $h+1$ 步仿真时的可行动作集合; $\Omega_{S,h}$ 为第 h 步仿真时能够满足所有约束条件且当前奖励优于不执行动作的动作集合; $\Omega_{T,h}$ 为第 h 步仿真时所有可行的拓扑动作集合; $\Omega_{R,h}$ 为第 h 步仿真时所有可行的发电出力调整动作集合; $\Omega_R(\cdot)$ 为发电出力调整动作集合; $F_h(a_i)$ 和 $F_{h,\max}$ 分别为第 h 步仿真时的惩罚函数及惩罚函数的最大值。

2.3 结合领域知识的学习策略

2.3.1 互为后备的双学习模型

在本文中,由于计算资源有限,所构建深度神经网络规模较小,因此无法保证在所有场景与数据集上都能收敛。为此,本文随机划分了数据集,并由A3C模型构建了2个不同的智能体,依据所划分的数据集分别对这2个模型进行训练。在此基础上,令2个智能体在在线应用阶段互为后备,当其中之一失效时由后备智能体求得拓扑或发电出力调整动作(若其中一个模型无法满足约束时,则另一个后备模型仍然有机会使系统恢复安全)。此外,本文使用2种不同的奖励函数训练2个智能体,从系统安全性、运行成本2个不同的角度优化拓扑控制决策。

2.3.2 奖励函数

在深度强化学习过程中,奖励函数用于评价智能体动作的效果,并引导智能体调节其参数使预期奖励最大化。本文采用2个智能体互为后备,其奖励函数分别从改善潮流分布均匀程度、降低系统运行成本2个角度进行设计。

首先,提高系统潮流分布的均匀程度有助于提高系统运行的安全性^[16,22]。基于输电元件实际输送容量与其最大输电能力之比,可以最大化以线路平均传输容量为目标的默认奖励函数 R_0 ,以反映电网运行均匀度。

$$R_0 = \sum_{l=1}^{n_l} \left[1 - \left(\frac{S_{L,l}}{S_{Lm,l}} \right)^2 \right] \quad (16)$$

式中: $S_{L,l}$ 为输电线路 l 的当前传输功率; $S_{Lm,l}$ 为输电线路 l 的传输功率额定最大值。

其次,后备模型以电网运行总成本最小为目标,奖励函数 R'_0 为正常数减去运行成本或停电损失。

$$R'_0 = \begin{cases} R_b - (c_1(t) + c_r(t)) & \text{正常运行} \\ R_b - c_b(t) & \text{停电} \end{cases} \quad (17)$$

式中: R_b 为正常数,作用为鼓励智能体尽可能运行更长时间,避免智能体因惩罚积累过多而使系统提前停止运行。

在基础奖励函数 R_0 和 R'_0 的基础上,若发生潮流越限,则在奖励函数上附加一个较大“惩罚”;若系统潮流发散,则给以智能体一个高额“惩罚”。此外,若惩罚项持续过高,预期累积奖励为负,则智能体可能会尝试直接使潮流发散以提前终止系统运行。为此,本文对系统运行时奖励最小值进行了约束。所使用的最终奖励函数 R 为:

$$R = \begin{cases} \max \left(-R_{\min}, R_0 - \sum_{l=1}^{n_l} \chi_l \max \left(0, \frac{S_{L,l}}{S_{Lm,l}} - 1 \right) \right) & \text{正常} \\ -R_g & \text{结束} \end{cases} \quad (18)$$

式中: R_{\min} 为系统运行时奖励最小值; R_g 为潮流发散时给予智能体的惩罚项; χ_l 为输电线路 l 潮流越限的惩罚系数。

2.3.3 强约束校验

若在训练过程中仅使用随机搜索,则因动作空间过大,难以在有限时间内获得足够高质量样本。为了提升样本质量,使强化学习智能体更快获得理想效果,本文在离线强化学习过程中强制进行潮流约束检验,其流程如图3所示。在智能体采取运行控制动作后,在计算奖励的同时计算系统所有潮流

约束。若智能体采取的动作会造成约束越界,则将各动作概率按降序排列,并选择动作概率次高的运行控制动作。以此类推,直到智能体发现能够满足约束条件的动作。若所有动作均无法满足约束条件,则智能体将执行最大化当前奖励函数的动作。

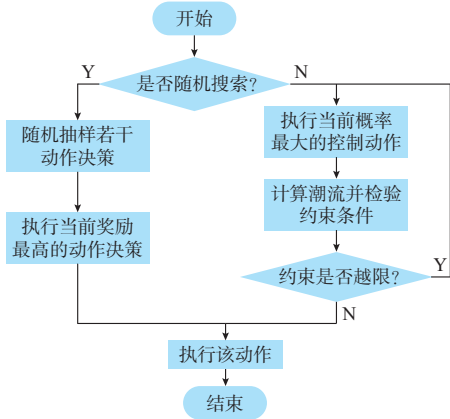


图3 结合强制约束校验的训练搜索流程

Fig. 3 Training searching process combining forced constraint verification

由于动作空间过于巨大,传统的深度强化学习难以有效探索遍历可能的动作,因此局部最优难以避免,且训练时间过于漫长。通过上述强制约束校验过程,可以显著降低搜索阶段的计算量,加快深度强化学习的训练速度。

3 算例分析

本文仿真验证在CPU内存为16 GB、GPU为GTX1070的计算机上进行。所用于仿真的Python模块主要包括Grid2op、Tensorflow、Keras、Pypowernet、Networkx。本文模型与测试代码均已开源^[23]。

3.1 环境介绍

本文仿真所用系统的详细信息及数据集场景来源详见附录A。该系统有35个变电站、22台发电机、59条输电线路。考虑不同的变电站母线连接方式,该系统共有177个节点,每步优化时有65 536种不同的拓扑结构决策。当线路潮流超越热稳定极限时,线路将直接断开。当线路潮流持续超过安全约束时,线路也将断开。该系统的拓扑结构如附录A图A1所示。

为了验证本文方法的有效性以及计算效率,采用无控制的参考方法以及基于穷举拓扑决策的混合整数优化(仅在约束越限时进行计算,搜索使系统恢复安全的拓扑结构,若发现系统能满足约束条件则应用该拓扑结构并停止搜索)方法进行对比研究。

需指出的是,65 536种不同拓扑结构决策并不能简单视为某连续决策变量的离散化,原问题极度非凸且复杂度高,无法使用主流求解器(如Cplex等)进行有效求解。文献中现有混合整数优化仅能考虑少量不同的拓扑决策^[2-16],无法实现本问题的求解。

3.2 仿真结果

本文仿真通过降低学习率并使智能体在更大的数据集上进行训练,同时在本地测试数据集上实现所有场景下的安全运行,如图4和图5所示。在图4和图5中,蓝色饼图表示系统安全运行的时长占总仿真时长的百分比^[23],括号内的数值是测试样本数量。由图4和图5可知,本文方法通过拓扑结构优化可有效实现系统的安全运行。图4实现了所有测试场景的安全运行,而所对比的穷举方法仍然难以实现所有场景下的安全运行,其在图5(c)场景中仅能在13.5%的时长内安全运行。

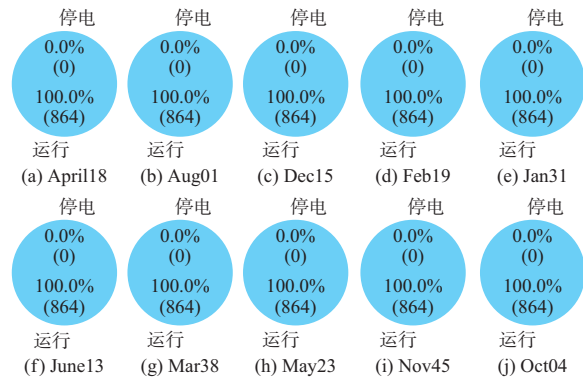


图4 采用深度强化学习智能体的仿真结果
Fig. 4 Simulation results with adoption of deep reinforcement learning agent

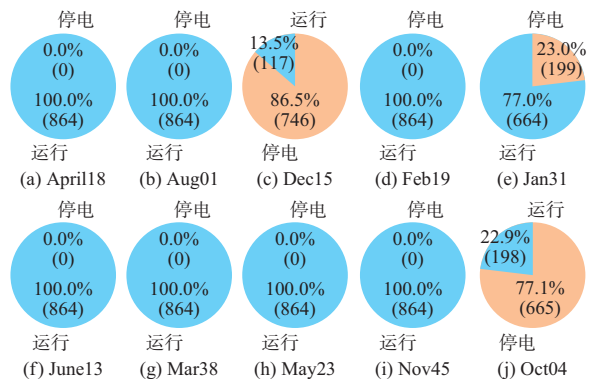


图5 采用穷举方法的仿真结果
Fig. 5 Simulation results with adoption of exhaustive method

为了验证本文方法的经济性,表1和表2对比了各方法的运行成本及其优化比例。附录A图A2对比了潮流未发散情况下各方法的系统运行成本。

从表2可看出,本文方法较无拓扑结构优化的对照方法降低了97.17%的总成本,较穷举方法提升更为明显。此外,在未出现停电的5个场景(April18、Aug01、June13、Mar38、May23)下,本文方法也通过改善潮流分布降低了运行成本。成本降低主要原因在于本文方法自适应地求解得到了各状态下最优的拓扑决策,能够使系统在各种随机因素作用下都能安全运行,并极大地降低停电损失。

表1 不同场景的运行成本对比
Table 1 Comparison of operation costs in different scenarios

运行场景	运行成本/万美元		
	本文方法	无控制	穷举方法
April18	36.37	36.45	56.28
Aug01	25.83	25.88	25.84
Dec15	56.28	4 276.00	42.60
Feb19	54.87	2 638.00	55.58
Jan31	43.14	3 782.00	3 451.00
June13	20.54	20.55	23.32
Mar38	47.11	47.22	47.20
May23	21.98	22.01	22.01
Nov45	54.87	2 638.00	55.58
Oct04	43.14	3 782.00	3 451.00

表2 总成本及优化比例对比
Table 2 Comparison of total costs and optimization proportion

方法	总停电成本/ 万美元	总运行成本/ 万美元	总成本/ 万美元	优化比例/ %
本文方法	0	473.8	473.8	97.17
无控制	16 500.0	221.2	16 720.0	0
穷举方法	8 723.0	385.3	9 108.0	45.53

表3对比了不同方法在不同仿真算例场景下的总计算时间。由表3可见,本文所提出的方法较穷举方法极大地降低了计算负担,穷举方法完成计算(仅在约束越限时穷举)的总计算时间为3.9 h,难以在实际系统运行决策间隔中投入使用,而本文方法在系统运行过程中进行了8 640次优化决策的总计算时间为0.3 h,节省了92.26%的计算时间,具有在线使用的潜力。

表3 总计算时间对比
Table 3 Comparison of total computation time

方法	计算时间/s	节约时间/s	时间缩减比例/%
本文方法	1 085.41	12 943.87	92.26
无控制	无	无	无
穷举方法	14 029.28	无	无

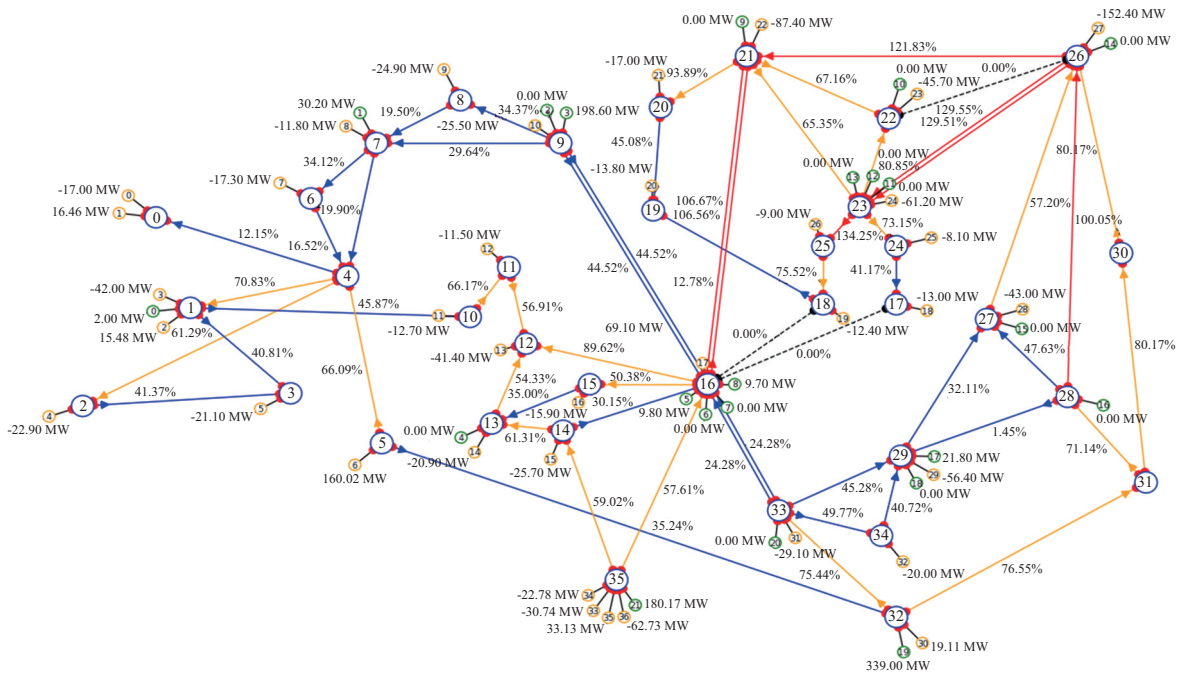
常规穷举方法在约束越限时须对不同的拓扑结构组合进行多次潮流计算,过程极其耗时,而本文方法可通过训练强化学习智能体,将大规模仿真的计算耗时转移至离线神经网络训练过程,使在线决策阶段快速求得满足系统安全的拓扑结构最优解。

为了验证智能体对于系统拓扑结构优化的实际效果,本文也在潮流发散前对系统在有控制情况下的线路潮流分布进行了对比,详见图6与附录A图A3。由图6可见,在2012-01-23T08:25:00时,无拓扑结构优化的参考算例在输电线路16-17等多条输电电路上出现了跳闸或停运,且输电线路16-23、输电线路26-23与输电线路23-25的载流量也超过了额定安全值。相较而言,本文所提出的拓扑结构优化方法通过定义奖励函数来惩罚潮流约束超限,引导智能体在环境中持续运行,可有效调整系统潮流分布,使线路潮流维持在约束范围内并避免连锁跳闸,从而改善系统运行的安全性。

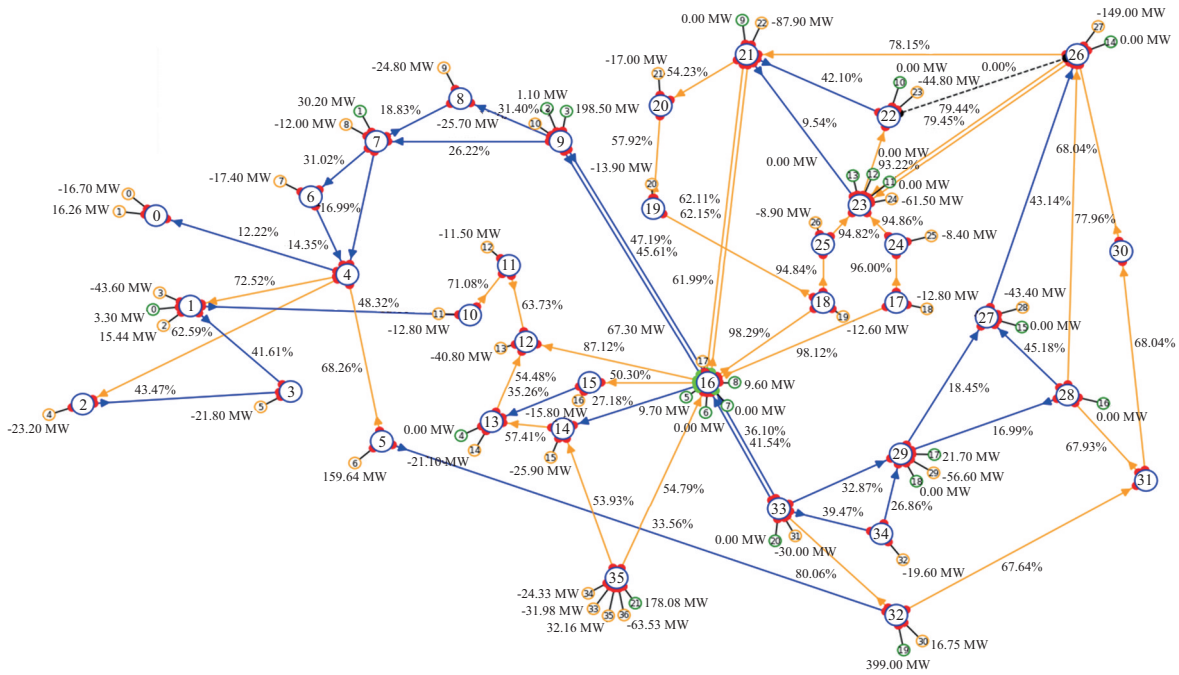
4 结语

本文提供了一种结合A3C深度强化学习与电力系统领域知识的系统拓扑与发电调整控制方法,以解决电力系统在线运行中决策变量计算量大与前瞻性不足的问题。为了在巨大的动作空间下高效地探索可行拓扑决策,本文引入A3C强化学习框架,使各个智能体分别与不同的环境相交互,增加了样本的多样性,克服了训练难以收敛的问题。同时,为了使智能体能有效地对电力系统状态进行建模,本文所设计智能体考虑了拓扑结构特征与数值特征,并以若干先验可行的动作作为动作空间。由于搜索量巨大,本文为了加快算法收敛并提高智能体性能,结合了电力系统领域知识设计奖励函数与强制约束校验。在一定条件下,智能体将强制寻找可以满足约束条件的动作,以避免过多无效的随机搜索。最后,通过随机训练数据集的划分,由不同数据训练的2个神经网络互为后备,提高了电力系统运行控制的安全性能。

本文训练主要基于处理后的系统数值特征,并不一定能充分反映各种情况下的网络拓扑特征与潮流的空间分布,后续研究可尝试由图神经网络建立智能体以提取潮流空间分布特征。此外,因该问题决策变量维数过高,搜索空间仍然巨大,后续研究可考虑以系统的图论目标状态为输出来改进动作集的建立过程。



(a) 无拓扑结构优化



(b) 本文智能体优化

— 输电线路; — 变电站出线; — 负荷; — 发电机; — 无母线; — 母线1; — 母线2; ○ 负荷节点; ○ 发电机节点; ○ 变电站

图6 系统潮流分布

Fig. 6 Power flow distribution of power system

附录见本刊网络版 (<http://www.aeps-info.com/aeps/ch/index.aspx>), 扫英文摘要后二维码可以阅读网络全文。

参考文献

[1] 杨知方, 钟海旺, 夏清, 等. 输电网结构优化问题研究综述和展

望[J]. 中国电机工程学报, 2016, 36(2): 426-434.

YANG Zhifang, ZHONG Haiwang, XIA Qing, et al. Review and prospect of transmission topology optimization [J]. Proceedings of the CSEE, 2016, 36(2): 426-434.

[2] HAN J, PAPAVALIOU A. The impacts of transmission topology control on the European electricity network[J]. IEEE Transactions on Power Systems, 2016, 31(1): 496-507.

- [3] 赵博石,胡泽春,宋永华.考虑 $N-1$ 安全约束的含可再生能源输电网络鲁棒优化[J].电力系统自动化,2019,43(4):16-24.
ZHAO Boshi, HU Zechun, SONG Yonghua. Robust optimization of transmission topology with renewable energy sources considering $N-1$ security constraint[J]. Automation of Electric Power Systems, 2019, 43(4): 16-24.
- [4] 蒋航,方颖颖,周建英,等.基于网络拓扑结构优化的电力系统网损在线综合优化方法[J].自动化与仪器仪表,2020(11):210-215.
JIANG Hang, FANG Yingying, ZHOU Jianying, et al. Online comprehensive optimization method for network loss of power system based on network topology adjustment[J]. Automation & Instrumentation, 2020(11): 210-215.
- [5] 李少岩,任乙沛,顾雪平,等.基于短路电流约束显式线性建模的输电网络结构优化[J].电工技术学报,2020,35(15):3292-3302.
LI Shaoyan, REN Yipei, GU Xueping, et al. Optimization of transmission network structure based on explicit linear modeling of short circuit current constraints [J]. Transactions of China Electrotechnical Society, 2020, 35(15): 3292-3302.
- [6] LAN T, WANG W Z, HUANG G M. Transmission grid topology control using critical switching flow based preventive stabilizing redispatch [J]. IEEE Transactions on Power Systems, 2018, 33(3): 2572-2582.
- [7] YANG Z F, ZHONG H W, XIA Q, et al. Optimal transmission switching with short-circuit current limitation constraints[J]. IEEE Transactions on Power Systems, 2016, 31(2): 1278-1288.
- [8] FISHER E B, O' NEILL R P, FERRIS M C. Optimal transmission switching [J]. IEEE Transactions on Power Systems, 2008, 23(3): 1346-1355.
- [9] FULLER J D, RAMASRA R, CHA A. Fast heuristics for transmission-line switching [J]. IEEE Transactions on Power Systems, 2012, 27(3): 1377-1386.
- [10] SOROUGH M, FULLER J D. Accuracies of optimal transmission switching heuristics based on DCOPF and ACOPF [J]. IEEE Transactions on Power Systems, 2014, 29(2): 924-932.
- [11] KHANABADI M, GHASEMI H, DOOSTIZADEH M. Optimal transmission switching considering voltage security and $N-1$ contingency analysis [J]. IEEE Transactions on Power Systems, 2013, 28(1): 542-550.
- [12] WU J, CHEUNG K W. Incorporating optimal transmission switching in day-ahead unit commitment and scheduling [C]// 2015 IEEE Power & Energy Society General Meeting, July 26-30, 2015, Denver, USA: 1-5.
- [13] GOLDIS E A, RUIZ P A, CARAMANIS M C, et al. Shift factor-based SCOPF topology control MIP formulations with substation configurations [J]. IEEE Transactions on Power Systems, 2017, 32(2): 1179-1190.
- [14] HEIDARIFAR M, GHASEMI H. A network topology optimization model based on substation and node-breaker modeling [J]. IEEE Transactions on Power Systems, 2016, 31(1): 247-255.
- [15] GOLDIS E A, RUIZ P A, CARAMANIS M C, et al. Shift factor-based SCOPF topology control MIP formulations with substation configurations [J]. IEEE Transactions on Power Systems, 2017, 32(2): 1179-1190.
- [16] 张尚,王涛,顾雪平,等.基于输电网络结构优化的电网安全均匀调度[J].电网技术,2017,41(9):2957-2963.
ZHANG Shang, WANG Tao, GU Xueping, et al. Power grid homogeneous dispatch based on transmission topology optimization [J]. Power System Technology, 2017, 41(9): 2957-2963.
- [17] XIAO R S, XIANG Y M, WANG L F, et al. Power system reliability evaluation incorporating dynamic thermal rating and network topology optimization [J]. IEEE Transactions on Power Systems, 2018, 33(6): 6000-6012.
- [18] ZHANG Z D, ZHANG D X, QIU R C. Deep reinforcement learning for power system applications: an overview [J]. CSEE Journal of Power and Energy Systems, 2019, 6(1): 213-225.
- [19] ZHAO T Q, WANG J H. Learning sequential distribution system restoration via graph-reinforcement learning [J/OL]. IEEE Transactions on Power Systems [2021-03-04]. <https://ieeexplore.ieee.org/document/9508140>.
- [20] YOON D, HONG S, LEE B J, et al. Winning the L2RPN challenge: power grid management via semi-Markov after state actor-critic [C]// International Conference on Learning Representations, May 4-7, 2021, Vienna, Austria: 1-6.
- [21] HOSSAIN R R, HUANG Q H, HUANG R K. Graph convolutional network-based topology embedded deep reinforcement learning for voltage stability control [J]. IEEE Transactions on Power Systems, 2021, 36(5): 4848-4851.
- [22] 孙伟卿,王承民,张焰,等.电力系统运行均匀性分析与评估[J].电工技术学报,2014,29(4):173-180.
SUN Weiqing, WANG Chengmin, ZHANG Yan, et al. Analysis and evaluation on power system operation homogeneity [J]. Transactions of China Electrotechnical Society, 2014, 29(4): 173-180.
- [23] A solution for learning to run the power network challenge [R/OL]. [2020-10-20]. https://github.com/ZM-Learn/L2RPN_WCCI_a_Solution.

严梓铭(1994—),男,博士研究生,主要研究方向:电网运行与控制。E-mail:yanzmics@gmail.com

徐岩(1985—),男,通信作者,博士,副教授,博士生导师,主要研究方向:电网稳定与控制、微网与智能电网中的数据分折。E-mail:xuyan@ntu.edu.sg

(编辑 顾晓荣)

Topology Optimization of Power Systems Combining Deep Reinforcement Learning and Domain Knowledge

YAN Ziming, XU Yan

(School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore)

Abstract: While the flexibility of power system operation can be improved by topology optimization, the dimension of system-level discrete decision variables, including the connections of lines and substation busbars, is prohibitively high. Thus, the topology optimization problem of power systems can hardly be solved by the conventional mixed-integer optimization method. Aiming at this problem, a reinforcement learning based method is proposed combining asynchronous advantage Actor-Critic (A3C) and power system domain knowledge, which transfers the computational burden of online optimization to the offline agent training stage. The defined reward function is adopted to minimize the violations of power transmission line flow limits. Forced constraints verification is employed to reduce the searching space and improve the efficiency of the reinforcement learning. The fast computation of the topology optimization of power system operation is realized, and the operation security of power systems is enhanced. The effectiveness of the proposed method is validated by simulation testing results.

Key words: power system operation; deep reinforcement learning; topology optimization; power output adjustment

