

基于深度强化学习的配电网负荷转供控制方法

王光华¹, 李晓影¹, 宋秉睿², 张沛²

(1. 国网河北省电力有限公司 保定供电分公司, 河北 保定 071000; 2. 天津相和电气科技有限公司, 天津 300042)

摘要:随着城市规模的快速扩张以及电能替代的不断推进,配电网节点数大量增加,结构愈加复杂,发生故障后拓扑变化不确定性较大,传统负荷转供方法难以在短时间内给出高质量的解决方案。为此,提出基于深度强化学习的配电网负荷转供控制方法。将负荷转供过程视为一个马尔可夫决策过程,与配电网实时电气、拓扑数据进行交互,对联络开关与分段开关进行控制。为了提高算法的精度与泛化能力,针对算法动作策略加入了预模拟机制,调整了动作与学习的比例并采用自适应优化算法进行求解。算例分析表明,所提方法能够应对不同故障下配电网的拓扑变化,即时给出负荷恢复量、电网损耗、开关动作次数多方面最优的转供控制方案,这对于减小故障后的停电损失与提高用户满意度有着重要意义。

关键词:深度强化学习;配电网;负荷转供;马尔可夫决策过程

中图分类号:TM 761

文献标志码:A

DOI:10.16081/j.epae.202204015

0 引言

配电网具有闭环设计、开环运行的特点,联络开关为常开开关,通常在运行情况下保持断开状态,用于紧急备用转移,分段开关为常闭开关,通常在运行情况下保持闭合状态,其主要功能是隔离故障。当采用辐射状结构运行的配电网发生故障时,故障线路下游的负荷都会中断供电,需要尽快地切除故障设备,并快速地恢复下游失电负荷用户。负荷转供通过联络开关和分段开关的操作与不同组合,改变供电的路径,在满足配电网开环方式运行、电压幅值和载流量等基本约束的前提下,将故障或检修停电范围内以及重载区域内的线路负荷转带至其余配电线路,从而达到降低故障或检修带来的失电损失和减少过载配电网线路的目的,提高配电网运行的可靠性、经济性与客户满意度^[1]。

随着城市规模的快速扩张以及电能替代的不断推进,各行各业对电力的需求迅速增长,配电网节点数大量增加,结构愈加复杂,配电网故障后拓扑变化不确定性较大,开关的组合与供电路径的选择出现了“组合爆炸”的问题。而且城市电网建设相对较为滞后,导致电网结构的不合理等问题比较突出,故障后转供裕度不足等问题更加大了负荷转供的难度^[2]。目前大多数配电网公司负荷转供主要采取的方法为人工经验决策方法,方案的优劣依赖于运行人员的经验水平,人为失误可能导致负荷损失^[3]。

目前,负荷转供的求解方法大致有以下几类。

①启发式算法有分层树搜索法^[4]、支路交换法^[5],利用启发式规则来减小搜索空间和降低问题的复杂度比较简单直观,计算过程中重复使用规则且进行多次仿真计算,解的优劣依赖于规则的制定与配电网初始状态,影响了速度与解的质量。②数学优化算法有多阶段优化法^[6]、非线性规划法^[7]、动态规划法^[8],将负荷转供转化为对数学问题的求解,但是当配电网结构庞大复杂、维数高时易出现“组合爆炸”的问题。③专家系统法^[9]能够调用已经保存的策略库,实时性好,适用性广,但是库的建立和集成费时费力,且当配电网建设发生变化时,专家系统需要不断进行调整。④随机搜索算法有禁忌算法^[10]、遗传算法^[11-12]、粒子群优化算法^[13],这些算法鲁棒性较好,但是搜索过程范围较大,需要进行大量的仿真计算,可能不收敛,算法的优化选取规则复杂,所用的时间较长。已有的研究方法都将负荷转供视为优化问题或者搜索问题来进行求解,尽管都采取了近似简化方法或缩小搜索空间的方式,但由于配电网拓扑存在较大不确定性,转供路径存在“组合爆炸”的问题,故障后进行大量计算无法满足故障恢复的时效性要求,只能牺牲求解质量来达到速度要求,难以解决求解质量与速度间的矛盾。目前,应用强化学习算法求解配电网负荷转供问题的研究仍为空白。负荷转供解决方案是由一系列开关动作构成,是一个连续的控制过程,而配电网的状态只取决于上一时刻的状态和动作,与历史状态无关,因此本文将负荷转供过程视为一个马尔可夫决策过程MDP(Markov Decision Process)^[14],应用强化学习RL(Reinforcement Learning)理论来处理这一问题^[15]。

为此,本文提出了基于深度强化学习DRL(Deep Reinforcement Learning)的配电网负荷转供控制方法,采用 Dueling 深度 Q 网络 DQN(Deep Q-

收稿日期:2021-05-29;修回日期:2022-02-22

在线出版日期:2022-04-14

基金项目:国网河北省电力有限公司科技项目(kj2021-014)

Project supported by the Science and Technology Project of State Grid Hebei Electric Power Co.,Ltd.(kj2021-014)

Network)算法,与配电网实时电气、拓扑数据进行交互,对联络开关与分段开关进行控制,自适应配电网拓扑变化的不确定性;针对算法动作策略加入了预模拟机制,调整了动作与学习的比例并采用自适应优化算法进行求解,提高了算法的收敛速度与鲁棒性,当配电网发生故障时可以即时给出高质量的转供决策方案。

1 负荷转供的强化学习模型

强化学习是一个与环境不断进行交互,获得反馈,更新策略,不断迭代直至学习到最优策略的过程。配电网作为强化学习的环境,在每一时刻 t ,可以为智能体提供当前配电网状态空间 S ,智能体分析决策的结果为开关动作 A ,将动作施加在环境上,环境发生的状态转移概率为 P ,环境反馈给智能体即时奖励值为 R 。强化学习智能体目标为通过有限的步数最大化累积奖励值,从而找到最优策略。负荷转供的强化学习模型如图1所示。图中: V 为电压向量; I 为电流向量; S_w 为配电网中支路的开关状态向量; F 为支路故障状态向量。

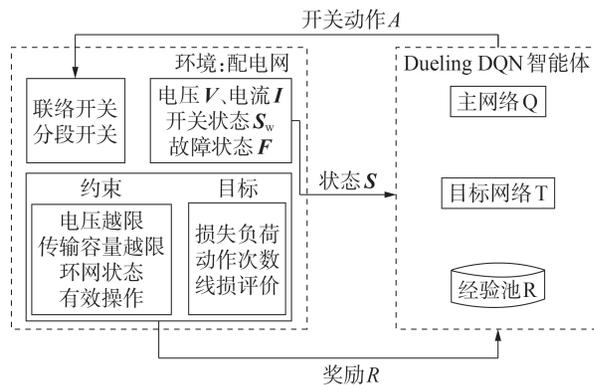


图1 负荷转供的强化学习模型

Fig.1 Reinforcement learning model for load transfer

1.1 状态空间

状态空间应尽可能考虑会对决策产生影响的因素。对于负荷转供问题,从数值角度出发,节点的电压、支路的电流反映了用户电压质量与线路载荷能力,是关键的分析数据;从空间角度出发,配电网的拓扑状态以及故障的位置信息可以作为选取合适的转供路径的依据。因此在强化学习中选择这些数据构建状态空间 S ,即:

$$S=[V, I, S_w, F] \quad (1)$$

1.2 动作空间与状态转移概率

一次完整的负荷转供操作由一系列联络开关与分段开关的投切所构成,为了防止状态空间过大,本文选取1次动作只投切1个开关的方式,所以相邻的2个状态之间,拓扑上的区别只有1个开关。此外负荷转供应该在有限的动作数内结束,应设置主动结

束本次转供的动作。所以智能体的动作空间选取为 $0, 1, \dots, 2N_s$,其中 N_s 为可操作的支路数量。当动作值为 $2N_s$ 时代表不采取任何操作并退出,本次决策结束,当动作值为 $0, 1, \dots, 2N_s-1$ 时,对 A 进行如下计算:

$$x=A \% 2 \quad (2)$$

$$y=\frac{A-x}{2}+1 \quad (3)$$

式中: x 为 A 除以2得到的余数; y 为动作支路的编号。由于每条线路用2个相邻的整数控制其投入或切除,用相邻奇偶位表示线路 y 的2种动作状态,具体如下:

$$x=\begin{cases} 0 & \text{投入编号为 } y \text{ 的支路} \\ 1 & \text{切除编号为 } y \text{ 的支路} \end{cases} \quad (4)$$

这样即可保证每次动作1条支路或者直接完成转供决策。在动作后若出现了电压越界、电流过载、动作次数超限等情况,则状态转移至失败退出状态;若恢复了全部非故障区域内的负荷,且配电网呈单辐射状,则状态转移至成功转供状态。由于有些情况下需要主动切除部分负荷以防止过载,因此智能体主动退出状态也会转移至结束状态。除此之外,任何其他状态都为过渡状态。

1.3 奖励函数

负荷转供首先要保证电网在约束范围内运行,以恢复用户供电,实现最佳的电能质量与经济效益为目标。本文将奖励函数分为奖励与惩罚2个部分,用于最终评价指导智能体的动作。

1) 奖励部分。

负荷转供的首要目标就是尽可能恢复所有用户供电,因此最重要的就是负荷恢复量。将动作累积恢复负荷量 R_L 作为目标函数的正比部分。

$$R_L = \sum_{k=1}^{N_{op}} P_{res,k} \quad (5)$$

式中: N_{op} 为总动作次数; $P_{res,k}$ 为第 k 次动作所恢复的负荷量。

转供应该在尽可能少的动作次数下完成,以降低运维的成本与失误的可能性,同时也节省操作的时间,防止配电网结构变化过大,为故障消除后恢复原运行方式增加难度。因此本文考虑将动作次数 N_{op} 作为目标函数的反比部分。

配电网的线损也是重要的成本因素,本文利用带电线路的阻抗参数,对配电网的线损情况进行评估,计算得到近似线损值 ΔP 为:

$$\Delta P = \sum_{i=1}^l I_i^2 r_i \quad (6)$$

式中: l 为带电线路总数; I_i, r_i 分别为流过第 i 条线路的电流和第 i 条线路的电阻。

目标函数 R_T 可根据各影响因素正反比关系计

算得到,同时为便于调节各因素对 R_T 的影响程度,以达到最佳训练效果,对各因素值分别添加偏置量 a_1, b_1, c_1 ,得到 R_T 的计算公式如下:

$$R_T = \left(\sum_{k=1}^{N_{op}} P_{res,k} + a_1 \right) \frac{1}{N_{op} + b_1} \frac{1}{\Delta P + c_1} \quad (7)$$

2) 惩罚部分。

维持正常的节点电压是保证配电网正常运行的基本要求,电压应保持在偏差为 $\pm 7\%$ 的容许范围内^[16],为了防止智能体动作时不满足约束条件,对于超出该范围的电压,予以高惩罚,对于范围内的动作不设置惩罚。

电压越限后状态应转移至失败退出状态,电压惩罚 $P_{V_{olt}}$ 计算方法如下:

$$P_{V_{olt}} = \begin{cases} P_U & U_i \leq U_{i,min}, U_i \geq U_{i,max} \\ 0 & U_{i,min} < U_i < U_{i,max} \end{cases} \quad (8)$$

式中: U_i 为节点 i 电压的标么值; P_U 为电压越限后的惩罚值; $U_{i,max}$ 、 $U_{i,min}$ 分别为节点 i 电压的上、下限,通常取1.07 p.u.和0.93 p.u.。

当传输容量超过线路与变压器的极限值时,易引发设备二次故障,本文取设备电流极限值作为运行上限,电流越限后状态转移至失败退出状态,电流惩罚 $P_{L_{im}}$ 计算方法如下:

$$P_{L_{im}} = \begin{cases} P_I & I_j \geq I_{j,max} \\ 0 & I_j < I_{j,max} \end{cases} \quad (9)$$

式中: P_I 为电流越限后的惩罚值; I_j 为设备 j 的电流值; $I_{j,max}$ 为设备 j 的电流值上限。

配电网正常运行时应呈辐射状,但也允许存在环网作为短时过渡状态,不允许作为长期运行状态出现,因此设置环网惩罚 $P_{L_{oop}}$ 时应分情况考虑,即:

$$P_{L_{oop}} = \begin{cases} P_f & g_n \in G_R \text{ (结束状态)} \\ P_m & g_n \in G_R \text{ (过渡状态)} \\ 0 & g_n \notin G_R \end{cases} \quad (10)$$

式中: P_f 为结束状态下存在环网的惩罚值; P_m 为过渡状态下存在环网的惩罚值; g_n 为已恢复的区域; G_R 为不包含分布式电源时所有配电网辐射状结构集合。

当智能体采取无效的操作,如对已经闭合的开关执行闭合动作,即采取重复操作,以及对故障打开线路进行动作时,该动作视为无效,给予无效动作惩罚 P_{Act} ,即:

$$P_{Act} = \begin{cases} P_A & a_k \in O_A \\ 0 & a_k \notin O_A \end{cases} \quad (11)$$

式中: P_A 为智能体采取无效动作时的惩罚值; a_k 为第 k 次的动作; O_A 为对故障打开线路进行动作的集合。

奖励函数中约束项 R_p 可通过求和得到,正常状态下有:

$$R_p = P_{V_{olt}} + P_{L_{im}} + P_{L_{oop}} + P_{Act} \quad (12)$$

最终模型的奖励函数 R 由奖励部分 R_T 与惩罚部分 R_p 构成,即:

$$R = R_T + R_p \quad (13)$$

2 基于深度学习的负荷转供算法

2.1 Dueling DQN 深度强化学习算法

Q学习是强化学习的主要算法之一,采用动作-价值函数 $Q(s, a)$ 来评估策略的优劣,即在某一状态 s 下,采取动作 a 能够获得奖励的期望,Q学习将状态和动作构建一张表来存储 Q 值,在不断更新学习中,根据 Q 值来选取能够获得最大收益的动作。

根据Bellman方程求解马尔可夫决策过程的最佳决策序列,在某状态下可能选择一系列动作构成最终决策 π ,状态值函数 $Q_\pi(s, a)$ 表明每个状态的值不仅由当前状态决定还由后续状态决定, $Q_\pi(s, a)$ 计算方法如下:

$$Q_\pi(s, a) = E_\pi \left[\sum_{i=0}^{+\infty} \gamma^i R_i \mid s_0 = s, a_0 = a \right] \quad (14)$$

式中: s_0 为当前状态; a_0 为当前状态下所执行的动作; γ 为折扣因子, $\gamma \in [0, 1]$ 表征未来回报相对于当前回报的重要程度; R_i 为第 i 次动作后获得的即时奖励; $E_\pi[\cdot]$ 表示针对策略 π 求 $[\cdot]$ 期望。

由于负荷转供过程由多次开关动作组成,每次动作的好坏不仅与当前状态有关,而且会影响未来的动作,进而影响最终转供方案的优劣。所以式(14)完全符合负荷转供的目标,即 $Q(s, a)$ 与之后做的所有可能的动作所获得的奖励都有关,即强化学习的目的是学习到整个负荷转供过程的最优控制策略。

负荷转供的强化学习模型是基于配电网运行数据进行决策的,其状态空间是由电压、电流等连续变量构成的连续空间,但Q学习采用 Q 值表来存储状态与动作的映射关系,在连续状态空间的情况下难以适用。因此本文采用深度强化学习,即DQN^[17]算法进行求解。DQN使用深度神经网络产生 Q 值将状态和动作相互映射,从大量训练中不断调整网络参数,在线寻求满足最大回报的最优控制策略。

DQN算法训练过程中在同一网络利用相同的 Q 值选择动作和评估动作,这种情况下DQN容易过高估计动作的 Q 值,过高估计的 Q 值易导致最终结果存在偏差,从而难以求得最优解。Double DQN^[18]对DQN算法的更新计算方式进行了改进,由于目标网络比主网络更新滞后,Double DQN不直接全部采用目标网络计算 Q 值,而在估计下一状态时先根据主网络选择动作,再用目标网络计算 Q 值,这样避免了盲目地过高估计 Q 值,提高了精确度,但其仍未考虑环境因素的影响,实际使用受到一定制约。

在配电网转供过程中,不同配电网状态下采取相同的动作可能带来完全不同的后果,为了能够同时考虑配电网环境因素的影响与动作带来的回报,本文采用考虑环境影响的 Dueling DQN 算法,将 Q 值函数分为配电网环境信息回报和动作回报,使学习的目标更明确,其神经网络结构图见附录 A 图 A1。

Dueling DQN 将状态向量作为输入,输出一个包含每个动作 Q 值的向量,神经网络中价值函数标量 $V(s_t)$ 仅与状态有关,与动作无关,其值表现了当前配电网状态的优劣;优势函数向量 $A(a_t)$ 进行了中心化处理,消除了配电网环境对开关动作的影响,体现了各开关动作之间的可辨识性,用来评估动作所带来的额外收益:

$$Q(s, a, \omega, \alpha, \beta) = V(s, \omega, \alpha) +$$

$$\left[A(s, a, \omega, \beta) - \frac{1}{|\mathcal{A}|} \sum_{a' \in \mathcal{A}} A(s, a', \omega, \beta) \right] \quad (15)$$

式中: ω 、 α 、 β 分别为公共隐藏层参数、价值函数层参数、优势函数层参数; \mathcal{A} 为所有动作的集合, $|\mathcal{A}|$ 表示集合 \mathcal{A} 中的元素个数,即状态 s 下的动作数; a' 为状态 s' 下具有最大 Q 值的动作,其中 s' 为状态 s 的下一状态。 \mathcal{A} 对向量 A 进行了中心化处理,突出了动作的差异,体现特定状态下各开关动作的优劣对比。

2.2 改进训练收敛效果与模型泛化

1) 预模拟-贪婪动作策略。

配电网负荷转供往往动作数量较多,且整个转供过程可能的动作组合极多,传统随机贪婪策略 (ϵ -greedy) 在前期选择动作随机探索时效率较低,在较大的动作空间内极易陷入局部最优动作,即使偶尔随机动作跳出局部最优,也难以影响学习的方向。且在选择最优动作时,神经网络对不常见的或复杂的故障状态缺乏泛化能力,可能选择次优的或完全错误的动作,这在实际应用中可能导致严重的后果。

为了解决这一问题,本文采用了改进的预模拟-贪婪动作策略,预模拟时首先采用集合对比的方式筛选排除无效/重复动作,该过程中无需进行仿真计算;其次保持原顺序取出前 k 个动作仿真预模拟,排除越限动作;然后分别按照负荷恢复量、线损值、 Q 值对动作进行三级排序;最后输出预模拟最优的动作作为真实动作。考虑到修改了算法原本的动作机制,可能导致算法收敛性出现问题,本文保留了一定比例的贪婪动作,具体如图 2 所示。

本文中预模拟-贪婪动作策略采用了一定数值的有限的预模拟动作数量,可以保证较大动作空间内的模拟仿真次数与时间。并且集合对比方法无需经过仿真计算即可排除大量无效动作,缩小了解空

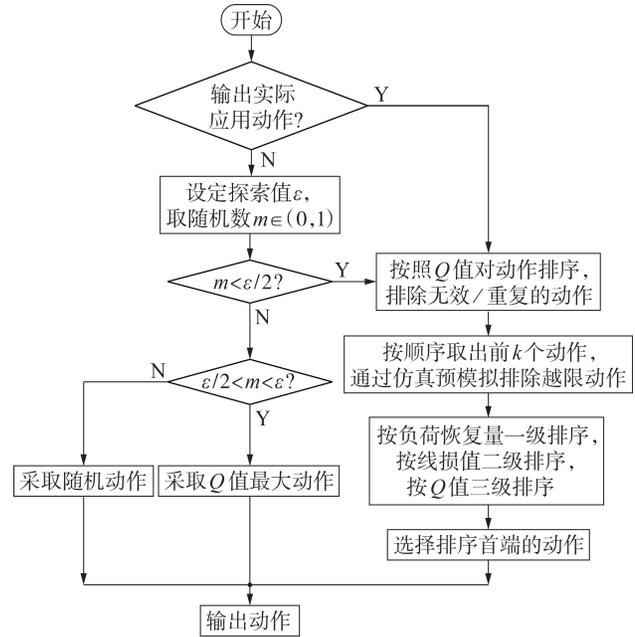


图 2 预模拟-贪婪动作策略

Fig.2 Pre-simulation-greedy action strategy

间范围,使强化学习训练时更容易找到最佳动作,加快了训练的速度;同时由于输入数据为大量配电网运行数据,失电时会引起大量输入数据的突变,变化过大的输入数据常常会导致神经网络学习振荡且难以收敛,保留一定程度贪婪动作策略中的动作机制能保证在波动数据下的训练效果与收敛性,使得神经网络能够清晰地辨别各状态下 Q 值最大的动作与真正的最优动作。

2) 调整智能体学习频率。

在常规 Dueling DQN 算法中,智能体每动作一次就要计算一次误差,更新学习一次,但实际上高频率的学习并不会提高神经网络的训练速度,在本文中高维度输入输出、神经网络参数庞大的情况下,反而会出现神经网络参数反复振荡的情况,导致参数难以收敛,学习速度较慢。因此本文调整了智能体学习频率,并在算例中确定最适合算法的比例关系,提高算法的学习能力。

3) Adadelta 优化算法。

在训练神经网络过程中,传统随机梯度下降法、Momentum 等优化算法采用固定学习率,收敛速度较慢且容易陷入局部最优解;自适应学习率优化算法 AdaGrad 对不同参数自动调节不同学习率,学习速度较快,但随着迭代次数增加学习率趋近于 0;本文采用的 Adadelta 优化算法可以加速神经网络的训练,无需设定学习率反复试错,而且避免了学习率越来越低的问题。

2.3 算法流程

基于深度强化学习的配电网负荷转供算法流程如图 3 所示,具体流程步骤见附录 B。

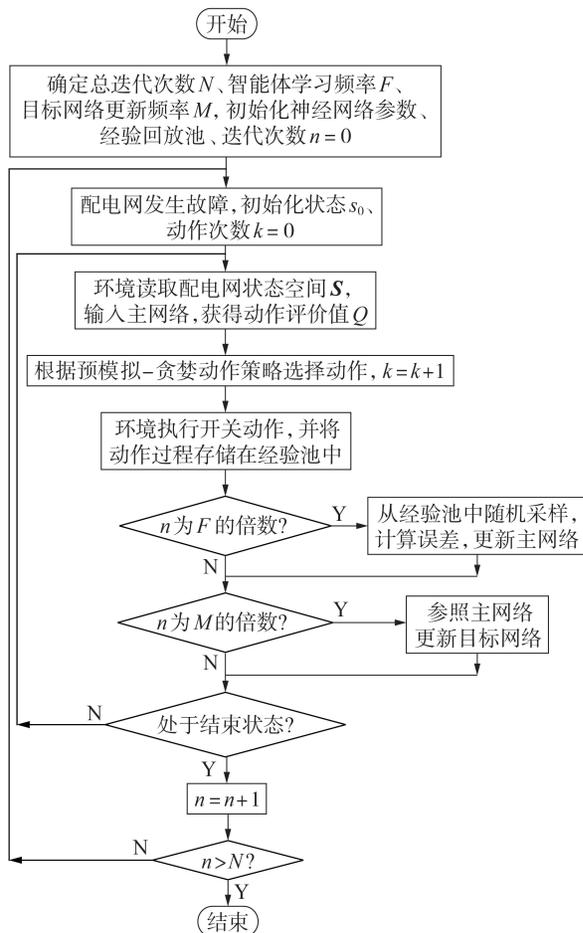


图 3 基于深度强化学习的配电网负荷转供方法流程图
Fig.3 Flowchart of load transfer method of distribution network based on deep reinforcement learning

3 算例分析

为了验证本文算法的有效性,采用 IEEE 33 节点配电系统进行验证,该配电系统包含 33 个节点以及 37 条支路,其中分段开关和联络开关分别有 32 个和 5 个。配电系统基准电压为 12.66 kV,总负荷为 3 826.06 kW+j2 366.49 kvar,基准功率为 10 MV·A,其拓扑结构如图 4 所示。

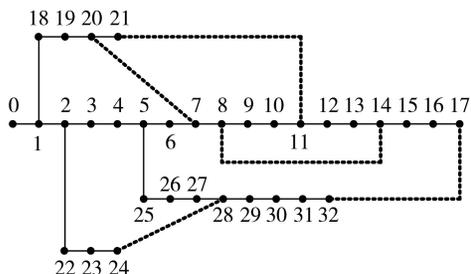


图 4 IEEE 33 节点配电系统结构
Fig.4 Structure of IEEE 33-bus distribution system

3.1 训练过程

利用 OpenDSS 仿真数据作为样本故障数据,在制造样本数据的过程中,采用了随机生成故障点的

方法。即在 32 个分段开关中随机选取 1 个开关发生故障并直接打开,其中各开关被选取的概率相同。5 条常开联络线由于作为备用线路,不发生任何故障。每回合开始动作之前,切除故障线路并进行仿真,确定配电网的初始状态,在此状态下,智能体选择操作故障线路将受到惩罚,并在状态转移至结束状态时结束此回合,自动生成新的随机故障,开始下一回合的负荷转供。

Dueling DQN 算法中公共隐藏层为 2 层,分别含有 512 和 256 个神经元,价值函数层与优势函数层均为 1 层,价值函数有 1 个神经元,输出为标量,优势函数有 75 个神经元,输出为与动作数相同的向量。激活函数均采用线性修正单元 ReLU,批处理数量为 200,折扣因子 γ 为 0.9,探索值 ϵ 取 0.95,经验池可存储 1 000 条经验数据,Adadelta 优化算法无需设定学习率,目标网络参数每 200 回合更新一次。

本文算法在 50 000 次动作回合后收敛,在转供决策时,动作的负荷恢复量等多方面接近最优,如图 5 所示。

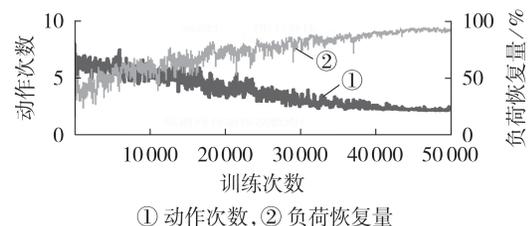


图 5 动作次数与负荷恢复量训练效果
Fig.5 Training effect of action times and load restoration

由图 5 可知,训练初期智能体对配电网转供环境不熟悉,初期动作次数较多且供电负荷比例低,转供控制效果较差;随着训练次数的不断增加,智能体不断与环境进行交互获得经验,学习到了用较少动作次数恢复更多负荷的方法,实现了负荷转供的最佳决策控制。

在共计 14 652 个回合的转供训练过程中,由于存在一定比例的探索性随机动作以及无法实现完全转供的情况,所以最终有 8 874 次转供控制实现了恢复全部负荷。图 6 为恢复全部负荷的情况下线损率随训练次数的变化情况,可见在后期智能体学习到了最佳运行成本的控制策略。

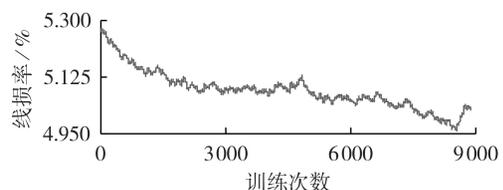


图 6 线损率训练效果
Fig.6 Training effect of line loss rate

3.2 负荷转供结果对比

对故障发生后的配电网的负荷转供进行测试。设支路12-13发生永久性故障并进行隔离后,故障点下游负荷点13—17为失电区域,可供转供直接使用的联络线有8-14、17-32,但这2条联络线单独对失电负荷转供时都会发生电压越下限的情况,必须采用多次动作转移负荷再进行转供,分别采用本文算法与传统强化学习算法、文献[19]中的启发式混合算法进行对比,其结果如表1所示,表中网损率由网络损耗除以网络总传输功率计算得到。

表1 负荷转供结果对比

Table 1 Comparison of load transfer results

算法	开关动作顺序	动作次数	负荷损失 / %	网损率 / %
本文算法	①合11-21,②合8-14, ③断6-7	3	0	4.05
传统强化学习算法	①合8-14,②断14-15	2	5.65	4.38
启发式混合算法	①合24-28,②合8-14, ③断5-25	3	0	4.45

从表1中可知,本文算法将重载线路上的部分负荷经过2条联络线转移至轻载线路上,在较少的动作次数内恢复全部负荷,且达到较低的网损水平;传统强化学习算法基本所有的尝试动作都会导致越限,所以最终切除部分负荷;启发式混合算法转移负荷后仍采用重载线路转供,恢复所有负荷后网损率比本文算法高出0.4%,由此可见本文算法基本实现给出了最佳控制策略。

对于不同故障下配电网的拓扑变化,即配电网状态信息发生变化,本文算法也可以即时给出转供方案,表2为采用本文算法时部分故障情况下的动作输出。可见在转供会引起电压越限时,智能体会切除较小的负荷以保证正常供电。在保证不越限的前提下,优先执行各方面综合最优的动作。

表2 不同故障下的动作输出

Table 2 Action output under different faults

故障线路	开关动作顺序	动作次数	负荷损失 / %	网损率 / %
4-5	①断5-6,②断16-17, ③合11-21,④合24-28	4	2.35	4.60
5-25	合24-28	1	0	4.56
10-11	①断27-28,②合11-21, ③合24-28	3	0	3.73

负荷转供方案需要多次进行动作,产生动作“组合爆炸”的问题导致实际求解的空间极大。本文算例中的37条支路,由于每条支路分为开、关2种状态,每次动作空间大小为74,在某故障下需要3次动作进行转供时其动作组合数为405 224个,需要5次动作进行转供时其动作组合数则达到约 2.2×10^9 个。

而对于不同的故障也需要求解不同的动作。本文算法在50 000次动作时训练趋于收敛,实现故障时能够从极大的动作组合中找到接近最优的动作策略并即时给出。

由于基于深度强化学习的配电网负荷转供控制方法具有离线学习、在线应用的特点,在线时无需在故障后花费大量时间仿真迭代,本文将单次动作的预模拟仿真数设置为5,即每次动作仿真5次,其单个动作时间基本为0.04 s,所以决策时间主要取决于动作次数,在本文算例中大部分故障动作次数都为1~5次,在线应用时求解过程的总仿真次数通常不超过25次,决策时间约为0.04~0.20 s,启发式混合算法与遗传算法在IEEE 33节点系统及更小的系统中决策时间分别为0.808 s及20.80 s^[20],本文算法决策时间远小于故障后其他算法的仿真计算时间。因此训练完成的深度强化学习负荷转供模型可以即时给出控制策略,可以在极短的时间内恢复供电,实现较高的经济效益。

3.3 训练效果对比分析

为验证本文提出的基于Dueling DQN的配电网负荷转供控制方法的有效性,将基于Dueling DQN算法的调度方法与DQN、Double DQN算法进行对比,对其设置相同的奖励函数,训练过程动作所获得的平均奖励值对比见附录C图C1。

为了提升动作奖励以及加快训练速度,采用了预模拟-贪婪动作策略的动作选取方式,改进前原强化学习模型在120 000次动作时平均奖励值仍然振荡,选择动作的效率低下,仍然存在大量动作,平均奖励为0.37,本文方法在60 000次动作后即趋于稳定收敛,基本不会出现无效动作,平均奖励接近0.8。而相对于预模拟完全取代最大Q值选择方式,在波动较小的输入数据训练样本上展现了较好的收敛效果,在波动较大的输入数据样本上后者可能会完全不收敛。因此,本文方法具有较好的收敛能力与鲁棒性,大幅降低了训练模型所需的计算量。

通过设置不同的比例在相同动作次数60 000次时对比其平均奖励的大小,如表3所示。可见其最大奖励出现在动作学习比例为5时,比例为4、5时的平均奖励非常接近,但比例为4时的学习次数较多,所以会花费更多的时间进行学习。因此确定动作与学习的比例为5,即每5次动作进行1次学习时,本文算法具有最好的学习效果。

表3 不同动作-学习比例比较

Table 3 Comparison of different action-learning ratios

比例	平均奖励	比例	平均奖励
2	0.551	6	0.690
4	0.780	10	0.720
5	0.790		

本文提出的预模拟-贪婪动作策略防止了训练

过程中预模拟动作屏蔽 Q 值最大动作,使 Q 值最大动作有更多实际验证的机会,避免了学习过程中对 Q 值的过高估计,结果显示预模拟-贪婪动作策略训练过程收敛趋势稳定,验证了本文算法的稳定性和收敛能力。全部采用预模拟动作与保留一定比例贪婪动作策略的损失函数对比见附录C图C2。

4 结论

本文利用数据驱动思维,将强化学习方法应用于负荷转供控制。直接分析电网运行环境信息,从中提取有效信息并构建强化学习模型,通过控制配电网开关进而实现负荷转供,自适应配电网拓扑变化的不确定性,对于不同故障、故障类型和运行方式的控制策略分析无需调整模型。

本文算法通过改进动作机制提高了决策收益,加快了算法的训练速度。相对于传统算法,其具有离线学习、在线应用的特点,将计算量转移至离线,通过大量的离线学习积累经验,在配电网发生故障后能够快速地进行线上计算,在极短时间内为运行人员提供有效的控制策略,减少停电损失并降低运行成本,这对于提高客户供电满意度有着重要意义。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 盛万兴,孟晓丽,宋晓辉. 智能配电网自愈控制基础[M]. 北京:中国电力出版社,2012:22-23.
- [2] 杨期余. 配电网网络[M]. 北京:中国电力出版社,1998:32.
- [3] 马梦茹. 考虑停电经济损失的配电网10千伏母线故障恢复重构研究[D]. 郑州:郑州大学,2018.
MA Mengru. Study on fault recovery and reconstruction of 10 kV bus in distribution network considering blackout economic loss[D]. Zhengzhou:Zhengzhou University,2018.
- [4] 徐爱军,陈谦,王朝明. 基于分层思想的配电网负荷转供算法[J]. 科学技术与工程,2013,13(10):2827-2831.
XU Aijun, CHEN Qian, WANG Chaoming. An efficient algorithm for distribution network service restoration[J]. Science Technology and Engineering,2013,13(10):2827-2831.
- [5] 张栋,张刘春,傅正财. 配电网重构的快速支路交换算法[J]. 电网技术,2005,29(9):82-85.
ZHANG Dong, ZHANG Liuchun, FU Zhengcai. A quick branch-exchange algorithm for reconfiguration of distribution networks[J]. Power System Technology,2005,29(9):82-85.
- [6] 章立宗,吴磊,蒋玮,等. 考虑二次转供的配电网负荷转供方案多目标优选技术[J]. 电网技术,2019,43(7):2321-2327.
ZHANG Lizong, WU Lei, JIANG Wei, et al. Research on distribution network load transfer scheme based on multi-objective optimization considering secondary re-sale strategy[J]. Power System Technology,2019,43(7):2321-2327.
- [7] 许寅,王颖,和敬涵,等. 多源协同的配电网多时段负荷恢复优化决策方法[J]. 电力系统自动化,2020,44(2):123-131.
XU Yin, WANG Ying, HE Jinghan, et al. Optimal decision-making method for multi-period load restoration in distribution network with coordination of multiple sources[J]. Automation of Electric Power Systems,2020,44(2):123-131.
- [8] XU Junbo, WANG Huiqiang, FENG Guangsheng. A novel service recovery method based upon Bellman dynamic programming[C]//International Conference on Computer and Information Application. Tianjin, China:IEEE,2010:110-115.
- [9] LIU C C, LEE S J, VENKATA S S. An expert system operational aid for restoration and loss reduction of distribution systems[J]. IEEE Transactions on Power Systems,1988,3(2):619-626.
- [10] 向小蓉,刘涤尘,向农,等. 基于并行禁忌搜索算法的配电网重构[J]. 电网技术,2012,36(8):100-105.
XIANG Xiaorong, LIU Dichen, XIANG Nong, et al. Distribution network reconfiguration based on parallel tabu search algorithm[J]. Power System Technology,2012,36(8):100-105.
- [11] 孙大雁,林济铿,袁龙,等. 配网负荷转供最优方案确定新方法[J]. 中国电力,2013,46(11):57-61.
SUN Dayan, LIN Jikeng, YUAN Long, et al. A new method for determining optimal load transfer scheme in distribution network[J]. Electric Power,2013,46(11):57-61.
- [12] 刘延华,李永飞,李林,等. 配电网重构遗传算法的个体编码综述[J]. 计算机应用与软件,2018,35(10):33-37.
LIU Yanhua, LI Yongfei, LI Lin, et al. Individual coding overview of distribution network reconfiguration genetic algorithm[J]. Computer Applications and Software,2018,35(10):33-37.
- [13] 徐岩,张荟,孙易洲. 基于变异粒子群算法的主动配电网故障恢复策略[J]. 电力自动化设备,2021,41(12):45-53.
XU Yan, ZHANG Yun, SUN Yizhou. Fault recovery strategy of active distribution network based on mutation particle swarm optimization algorithm[J]. Electric Power Automation Equipment,2021,41(12):45-53.
- [14] 王雪松,朱美强,程玉虎. 强化学习原理及其应用[M]. 北京:科学出版社,2014:41-42.
- [15] 江昌旭. 基于分层协作强化学习的配电网负荷建模研究[D]. 广州:华南理工大学,2020.
JIANG Changxu. Load modeling of distribution network based on hierarchical and cooperative reinforcement learning[D]. Guangzhou:South China University of Technology,2020.
- [16] 国家技术监督局. 电能质量供电电压允许偏差:GB 12325—1990[S]. 北京:中国电力出版社,1990.
- [17] CHUNG J. Playing Atari with deep reinforcement learning[J]. Computer Ence,2013,21:351-362.
- [18] HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q-Learning[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Palo Alto, USA:AAAI,2016:2094-2100.
- [19] 袁龙. 电力系统配电网负荷转供研究[D]. 天津:天津大学,2012.
YUAN Long. The research of load transfer in distribution network[D]. Tianjin:Tianjin University,2012.
- [20] 黄程程,顾洁,方陈. 基于无向生成树的并行遗传算法在配电网重构中的应用[J]. 电力系统自动化,2015,39(14):89-96.
HUANG Hongcheng, GU Jie, FANG Chen. Application of undirected spanning tree-based parallel genetic algorithm in distributed network reconfiguration[J]. Automation of Electric Power Systems,2015,39(14):89-96.

作者简介:



王光华

王光华(1978—),男,高级工程师,主要研究方向为电力系统及其自动化技术(E-mail:13582994117@139.com);

李晓影(1982—),女,高级工程师,硕士,主要研究方向为电力系统及其自动化技术(E-mail:297783167qq.com);

张沛(1972—),男,教授,博士研究生导师,通信作者,主要研究方向为电力系统规划和运行(E-mail:peizhang166@qq.com)。

(编辑 李玮)

Load transfer control method of distribution network based on deep reinforcement learning

WANG Guanghua¹, LI Xiaoying¹, SONG Bingrui², ZHANG Pei²

(1. Baoding Power Supply Company, State Grid Hebei Electric Power Co., Ltd., Baoding 071000, China;
2. Tianjin Xianghe Electric Co., Ltd., Tianjin 300042, China)

Abstract: With the rapid expansion of urban scale and the continuous advancement of electric energy substitution, the number of buses in distribution network increases greatly, the structure becomes more complex, the uncertainty of topology changing after fault becomes larger, and it is difficult for the traditional load transfer methods to provide high-quality solution schemes in short time. For that, the load transfer control method of distribution network based on deep reinforcement learning is proposed. The load transfer process is regarded as a Markov decision process, which interacts with the real-time electrical and topological data of distribution network, and controls the tie switch and subsection switch. In order to improve the accuracy and generalization ability of the algorithm, the pre-simulation mechanism is added to the action strategy for the algorithm, the proportions of action and learning are adjusted and the adaptive optimization algorithm is adopted to solve the problem. The case analysis shows that the proposed method can deal with the topology variation of distribution network under different faults, and provide the optimal transfer control scheme of load recovery amount, power grid loss and times of switching operation immediately, which is of great significance for reducing the outage loss and improving users' satisfaction.

Key words: deep reinforcement learning; distribution network; load transfer; Markov decision process

(上接第211页 continued from page 211)

Real-time dispatching strategy of cascaded hydropower stations under frequent load fluctuation

YANG Yuqi^{1,2}, MO Li^{1,2}, ZHOU Jianzhong^{1,2}, XIAO Wenjing^{1,2}, XU Zhanxing^{1,2}

(1. School of Civil and Hydropower Engineering, Huazhong University of Science and Technology, Wuhan 430074, China;
2. Hubei Key Laboratory of Digital Valley Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: Under the environment of grid-connection of high proportion of renewable energy, how to utilize the regulation effect of hydropower safely and efficiently is one of the key issues for the safe and stable operation of power grid. The frequent load fluctuation situation faced by cascaded hydropower stations in the period of power system transformation is introduced, the aftereffect influence of real-time dispatching for hydropower stations is analyzed, and the quantitative criterions of hydropower flexibility in real-time scale is established. On this basis, a real-time dispatching strategy of hydropower stations under the situation of frequent load fluctuation is proposed, which adopts the mode of combining real-time and long-term perspectives and dynamically controlling the output of each period and the water level at the end of dispatching period at the same time to meet the demand of real-time adjustment. The case results show that the proposed real-time dispatching strategy can enhance the hedging ability of hydropower system against the risk of future dispatching period under multiple load fluctuation scenarios, and it is more adaptable for the current hydropower operation environment compared with the traditional real-time dispatching strategy.

Key words: load fluctuation; cascaded hydropower station; real-time dispatching; aftereffect; dispatching risk

附录 A

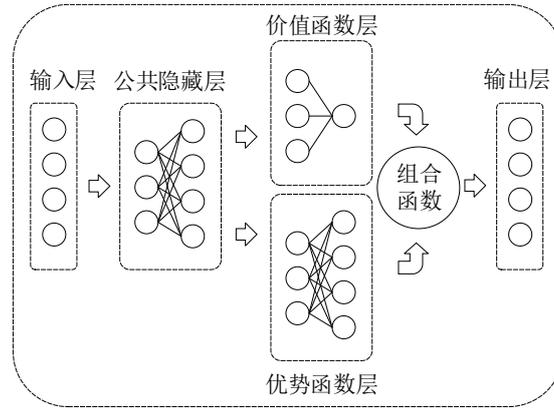


图 A1 Dueling-DQN 神经网络结构

Fig.A1 Neural network structure of Dueling-DQN

附录 B

Step1: 确定迭代次数 N 、学习比例 P 、目标网络更新频率 M ，初始化神经网络参数、经验回放池、总动作数 $n = 0$ 。

Step2: 配电网发生故障，开始负荷转供控制。初始化配电网状态即当前故障状态作为 s_0 ，动作次数 $k = 0$ 。

Step3: 环境读取配电网状态信息 S ，将其输入主网络，获得动作评价值 Q 。

Step4: 根据“预模拟-贪婪动作策略”选择相应动作。

Step5: 环境执行开关动作，并将动作过程存储在经验池中。

Step6: 每经过 P 次动作，从经验池中随机采样，计算误差，更新主网络。

Step7: 每经过 M 次动作，参照主网络更新目标网络。

Step8: 判断是否处于结束状态，若不是则从 Step3 开始继续动作，若是本回合转供结束，转供成功则给出转供控制策略。

Step9: 判断总动作次数是否达到设定的迭代次数，若不是则开始下一次转供，若是则结束训练，输出神经网络参数。

附录 C

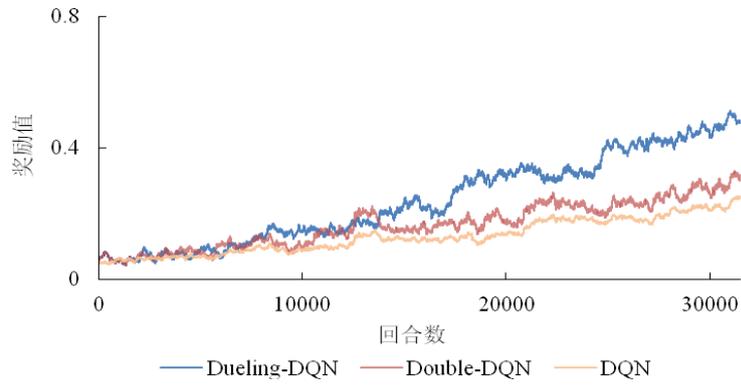


图 C1 不同算法平均训练奖励对比

Fig.C1 Comparison of average reward of different algorithms

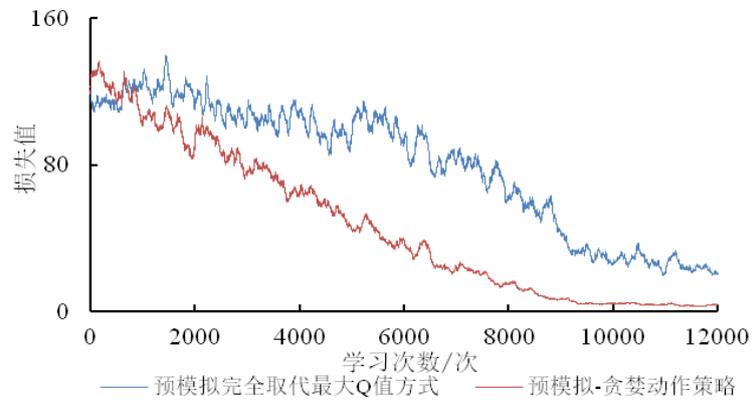


图 C2 不同算法平均损失对比

Fig.C2 Comparison of average loss of different algorithms