

DOI: 10.12158/j.2096-3203.2020.02.002

基于智能体群组强化学习的电网无功电压调控方法

范士雄¹, 刘幸蔚¹, 魏智慧², 刘瑞叶², 王松岩², 于继来²

(1. 中国电力科学研究院有限公司, 北京 100192;

2. 哈尔滨工业大学电气工程及自动化学院, 黑龙江 哈尔滨 150001)

摘要:无功电压调控是保障电网区域电压稳定的重要措施, 现有基于强化学习的单一智能体设计方式存在架构设计不合理等问题。文中提出了一种考虑节点电压幅值与电容器投切状况的电网综合运行状态描述方法, 设计了面向电网无功电压调控的强化学习智能体群组架构。该群组根据电网当前综合运行状态确定智能体成员, 并给出相应的无功调控动作, 各智能体成员将相邻时段电网状态改善程度作为奖励机制。算例表明该方法可应用于电网无功电压调控环境, 相较于单一智能体设计方式, 可以有效减少动作集数量, 更好地应对各种无功电压调控工况。

关键词:无功电压; 强化学习; 人工智能; 智能体群组; 电容器投切

中图分类号: TM761

文献标志码: A

文章编号: 2096-3203(2020)02-0010-08

0 引言

电网电压与无功关系密切, 局部无功实时平衡是保证全网电压稳定的前提条件, 因此电网必须要有足够的无功电源及无功补偿装置来满足全网的无功损耗和负荷的无功需求^[1-6]。深入研究电网动态无功电压调控方法对保障用户安全用电和电网可靠运行意义重大。

无功电压调控问题属于电力系统经典研究领域, 国内外学者对此展开了大量研究。经典优化算法受系统时变性、网络参数变化和电网运行条件等多因素影响, 易陷入局部最优解^[7-10]。文献[11—12]采用现代启发式算法求解动态无功电压调控的模型, 摆脱了非智能算法易陷入局部最优解的困局, 但该方法对建模的精确度有较高要求。无功电压调控是一个多变量、多约束、非线性复杂控制问题, 而当前人工智能技术的发展为解决该问题提供了新的技术途径。新一代人工智能算法可自主学习获取知识, 在解决非线性多约束动态规划问题上具有优势^[13-15]。文献[16]基于强化学习理论, 在不考虑负荷的时序变化的条件下, 在线给出当前时间断面的地区电网无功电压控制策略。文献[17]进一步将强化学习和迁移学习结合, 使用迁移学习将已有经验赋值给知识矩阵, 通过多个智能体协同合作更新知识矩阵, 加快了无功电压调控的寻优速度。文献[18]为克服控制变量增多造成的“维数灾”, 将大知识矩阵 Q 分块成小规模的知识矩阵块

链, 另外结合帝国主义竞争算法思想, 通过多个帝国主义间博弈, 协同搜索优化过程, 最终求得大规模电力系统的碳能复合流无功优化问题的最优解。

上述文献为利用人工智能技术解决无功电压调控问题提供了新思路。然而, 电网运行状态会跟随负荷变化而时序变化。面对电网中数量繁多的无功电压设备, 单一智能体的动作集为全网无功调控设备的动作组合, 动作集繁杂。因此, 无功电压调控不宜采用单一智能体模式。此外, 变电站当前时段投切电容器的动作不仅会改变电网当前的状态, 同时也会改变下一时段投切电容器的动作。这意味着电网时序动态无功电压调控是一个状态与动作高度耦合的强化学习问题。

基于强化学习的电网无功电压调控具有其特殊性与复杂性, 文中提出一种基于智能体群组强化学习的电网无功电压调控方法, 设计了包含多个智能体成员的智能体群组, 在时序场景下通过训练实现不同智能体成员的 Q 表收敛, 从而赋予群组内各智能体成员根据电网实时状态有效给出合理动作的能力。为简化描述, 文中无功电压调控只考虑投切电容, 其他无功电压调节方式(调变压器分接头等)可以在文中模型基础上拓展。

1 无功电压调控单一智能体设计方法

1.1 Q学习算法在无功电压调控中的应用

对于一个经典的无功优化问题, 假设系统中有 N 个节点, 其中 r 个节点变电站装有可投切电容器组。一般可建立如下无功优化模型:

$$\min f = \min \sum_{t=1}^{24} \sum_{i=1}^N |U_i(t) - U_i^{\text{spec}}| \quad (1)$$

收稿日期: 2019-09-23; 修回日期: 2019-10-18

基金项目: 国家电网有限公司科技项目“基于人工智能技术的电网调控体系架构及电网无功电压控制典型应用技术研究”

$$\text{s.t.} \begin{cases} g(x, u) = 0 \\ h(x, u) \leq 0 \end{cases} \quad (2)$$

式中: $U_i(t)$ 为 t 时刻节点 i 的电压幅值; U_i^{spec} 为调控期望值; g 为等式约束; h 为不等式约束; x 为节点状态量; u 为控制量。该无功电压调控优化模型以节点电压偏移量最小为目标。

由上述模型可见,无功电压调控追求系统节点电压达标。模型的等式约束为各时段潮流约束,不等式约束为发电机出力上下限约束及电容器投切约束等。因此从模型角度而言,无功电压调控是一个经典的多约束、非线性、含离散变量的非凸优化问题,求解难度极大。且在电网实际运行中,电网的线路及设备参数受外部环境的影响会随时发生变化,进一步对模型准确性造成不同程度影响。在此背景下,具有去模型化特点的强化学习技术为解决电网无功调控问题提供了另一条技术途径。

Q 学习算法属于强化学习算法中的主流分支算法。智能体通过探索未知环境获得代表经验的奖励,以 Q 值函数记录探索过程中连续动作的累积奖励,储存在 Q 表中,智能体依据 Q 表,选择下一个状态 Q 值最大的动作采取行动。算法采取动作的核心策略是“探索-利用”(“ ϵ -greedy”)策略,可有效避免决策陷入局部最优^[19-22],其数学描述如下:

$$a_{k+1} = \begin{cases} \arg \max_{a \in A} Q_{k+1}(s_{k+1}, a) & \epsilon > \epsilon_0 \\ a_s & \epsilon \leq \epsilon_0 \end{cases} \quad (3)$$

式中: ϵ 为服从均匀分布的随机数, ϵ 大于 ϵ_0 时,智能体按下下一个状态最大的 Q 值选择动作,否则,随机选择当前状态下的可执行动作。 a_s 表示按 0~1 范围随机从动作集中选择动作。

Q 学习算法用 Q 表存储状态-动作对应的 Q 值函数,其更新策略见式(4)。

$$Q(S, A) \leftarrow Q(S, A) + \alpha [R + \gamma \max_a Q(S', A) - Q(S, A)] \quad (4)$$

式中: A 为动作集; S 为状态集; R 为立即回报的奖励值; γ 为衰减率, $\gamma \in [0, 1]$; α 为学习因子, $\alpha \in [0, 1]$ 。由式(4)可见,当前 Q 值由奖励和后续环境状态下最优动作的 Q 值组成,奖励起到联系后续 Q 值与当前 Q 值的纽带作用,因此奖励的设计对 Q 值的更新至关重要。

在无功电压调控问题中,智能体根据电网的无功电压等运行状态特征,随机产生动作并根据动作结果产生的奖励值生成 Q 表。通过大量的离线学习与在线修正相配合实现 Q 表收敛后,智能体将通过 Q 表实现状态与最优动作的映射。由此可见,基于 Q 学习生成的智能体可有效回避无功电压调控

模型求解技术路线中的相关问题,具有技术潜力。然而若真正实现将 Q 学习技术有效应用于电网无功电压调控环境,需面对智能体的“状态”、“动作”及“奖励”环节与无功调控特征相结合的问题。

1.2 状态与动作的耦合

深入分析式(2),可发现对于不等式约束,除发电机出力有功与无功上下限约束等与时段无关的参数外,其他约束如电容器投切约束等均存在前后时段的相互影响。文中通过实例进一步详述该过程。假设电网中变电站装有电容器组,如图1所示。

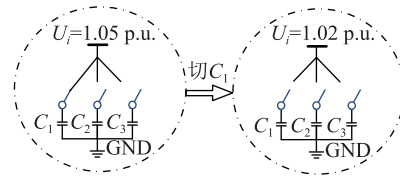


图1 节点 i 在 t 时刻的状态

Fig.1 State of the bus i at t

对多种投切动作组合分析后,该变电站可能的无功调控动作可归纳为以下3种情况:(1)变电站所有电容器均未投,变电站可执行动作为“投”或“保持”;(2)变电站所有电容器均已投,变电站可执行动作为“切”或“保持”;(3)变电站部分电容器投入运行,且仍有部分电容器可投,变电站可执行动作为“投”、“切”或“保持”。

对于装有3个可投切电容器组的节点,即图1,假设在 t 时段节点 i 的初态电压 U_i 为 1.05 p.u.,电压过高,此时电容器 C_1 接入电网, C_2 和 C_3 未投,则变电站当前时刻可以采取的电容器动作为“投”、“保持”或“切”。为改善节点 i 的高电压,变电站执行切掉电容器 C_1 的动作, $t+1$ 时段 U_i 降为 1.02 p.u.。此时变电站电压幅值仍较高,然而此时变电站所有电容器已切除,其可选动作只有“投”或“保持”,不能再“切”,因此 $t+1$ 时段变电站动作集与 t 时段存在差别。由此可见, t 时段动作理应作为 $t+1$ 时段电网状态描述的一个重要元素,因其直接影响了 $t+1$ 时段的动作集合。

由以上分析可见,经典数学模型中所存在的电容器投切约束存在前后时段的相互影响问题,反映在强化学习智能体设计中,就是动作与状态的耦合问题。因此从强化学习设计智能体角度而言,电网状态应由当前投入的电容器数量与节点电压幅值共同决定。若 t 时刻与 $t+1$ 时刻智能体动作集不一致,则2个时段应采用不同的智能体。

1.3 单一智能体模式的动作集

对于大电网而言,假设该电网有 K 个节点安装

无功调控装置。若采用传统的设计面向全网的单一智能体这一思路,每个节点电容器有“投”、“切”与“保持”3种动作(不考虑状态动作耦合),则对于 t 时段的电网状态,智能体将有 3^k 种动作组合(若 K 值为10,则动作组合有近6万种)。显然,如此大量的动作组合必将极大增加 Q 表收敛难度。此外,无功电压调控具有较强的局部调控特点,某个节点电压变化一般仅对电气距离较近的无功调控设备敏感。对于某个节点电压幅值较低的工况,单一智能体的大多数动作为无效动作,不利于智能体训练。

若令每个节点仅对应一个或少数智能体,则每个智能体的动作集将仅由该节点的少量无功调控设备决定,而非全网所有调控设备,从而有效降低智能体维度。从动作集设计角度来看,每个智能体成员的动作可视为全网单一智能体动作集分解后的子集。且多智能体设计思路还可以与无功电压的局部调控特征相契合,兼具强化学习与电气规律的双重合理性。

按智能体群组设计思路,对于某个待调控的电网状态,全网单一智能体在 $t+1$ 时段给出的复杂动作,将变成 N 个智能体成员在 $t+1$ 时段至 $t+N$ 时段给出各自相应动作。由于电网无功负荷变化比有功平缓且无功调控动作数量一般较少(即 N 值较小),将单时段复杂动作调控分解为连续多时段简单动作调控后,一般不存在因调控时间过长而导致的调控失效问题。

1.4 目标偏差奖励

一般而言,在单一智能体设计模式下需提前设计一个目标 f ,如式(1)。智能体依据动作效果与最优目标 f 两者间的偏差程度来判断动作效果的优劣并给出奖励值,以期获取最优动作。

虽然目标偏差奖励常用于无功电压调控的奖励设计,但该奖励方式在极端运行方式下无法获取最优动作。原因为电网在重载等特殊运行方式下存在某些节点电压过低工况,此时很可能无论智能体采取何种动作,其无功调控效果均与提前设定的 f 值偏差较大,导致所有动作的奖励均为负值,此时基于目标偏差奖励的智能体无法得到极端工况下的最优动作。

分析发现,即使动作集中所有动作均无法得到正向奖励,这些动作中仍存在一个最优动作,只是该最优动作的效果仍无法达到获得正向奖励的要求。由此可见,提前设定一个固定的、在某些工况下电网可能无论采用何种调控措施都无法达到的目标 f 值,智能体必然无法筛选出最优动作。

1.5 面向电网无功电压调控智能体设计需求

基于前文分析,面向电网无功电压调控的智能体设计应存在以下特征:

(1) 电网无功电压调控的状态描述应由当前无功调控设备状态与节点电压幅值状态共同决定;

(2) 面向电网无功电压调控应采取智能体群组方式。智能体群组包含多个智能体子群组,每个具有无功调控设备的变电站均有相应的智能体子群组。各子群组又由多个具有不同动作集的智能体组成,每个智能体成员只执行少数无功调控动作;

(3) 智能体成员奖励设计应采用状态改善奖励机制。状态改善奖励机制基于 t 时段当前状态与采取动作后下一个状态的比较,即使动作后系统仍未满足相应指标要求,只要系统 $t+1$ 时段状态较 t 时段有改善就会有正值奖励。

2 智能体群组设计方法

2.1 智能体群组设计方案

智能体群组参与电网无功电压调控过程如图2所示,群组以“电容器投入数量”作为智能体成员的选择判据。首先根据电压待调控节点已投电容器数量,群组选取相应的智能体成员;然后,该智能体成员根据目标节点电压幅值,给出调控动作;最后,基于状态改善奖励机制,将该动作调控效果与上一时段效果进行比较得到奖励值。反复此过程,各智能体成员形成各自的无功电压调控策略。

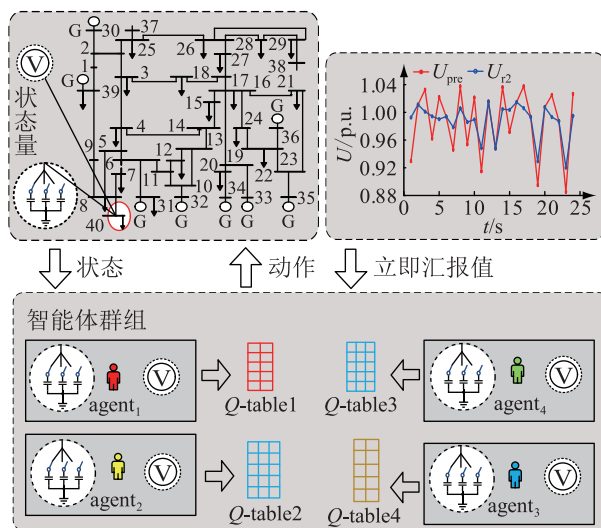


图2 智能体知识学习

Fig.2 Agent knowledge learning

2.2 电网无功电压环境状态集描述

文中电网的无功电压环境状态描述包括电压幅值和电容器投入数量两方面。考虑到电网负荷,电压幅值要求在 $0.95 \sim 1.05$ p.u.之间,按 0.1 p.u.进

行等间距划分,无论电压幅值低于 0.95 p.u. 还是高于 1.05 p.u. 均为越限状态。若目标节点已投入 n 个电容器,则其状态描述如表 1 所示。表中“ ∞ ”仅用于数学描述,不代表电网真实电压状态。

表 1 变电站投入 n 个电容器后电压状态集

Table 1 Environment state set with n integrated capacitors p.u.

状态名称	电压范围	状态名称	电压范围
1	$(-\infty, 0.95]$	7	$(1.00, 1.01]$
2	$(0.95, 0.96]$	8	$(1.01, 1.02]$
3	$(0.96, 0.97]$	9	$(1.02, 1.03]$
4	$(0.97, 0.98]$	10	$(1.03, 1.04]$
5	$(0.98, 0.99]$	11	$(1.04, 1.05]$
6	$(0.99, 1.00]$	12	$[1.05, +\infty)$

2.3 智能体成员 Q 表设计

若变电站节点有 N 个电容器,则投入电容器数量有 $N+1$ 种情况,该变电站对映的智能体子群组有 $N+1$ 个智能体。不同智能体针对变电站电容器组不同的运行状态采取动作,其动作集具有差异性。具体为:(1) 电容器全部投入时,动作集为“切”及“保持”;(2) 电容器全部切出时,动作集为“投”及“保持”;(3) 其他条件下,动作集为“投”、“切”及“保持”。

由于 Q 表的维数由状态集和动作集的元素数量决定,因此状态集和动作集确定后智能体成员 Q 表的维数也就随之确定。文中针对无功补偿设备的运行状态,直接将“电容器投入数量”作为智能体群组选择其成员的依据。综合电压幅值及投入电容器数量,智能体各成员及动作集设计如表 2 所示。

表 2 智能体群组中各成员及其动作集

Table 2 Agent group and its action set

智能体成员	电容器接入数量	动作集	Q 表维数
agent ₁	0	投、保持	12×2
agent ₂	1	投、保持、切	12×3
agent ₃	2	投、保持、切	12×3
...
agent _{N+1}	N	保持、切	12×2

可见,以往的强化学习智能体动作集一般仅设计 2 种动作^[15-16],文中在动作集中增加了“保持”,进一步完善了无功调控动作集。此外,文中智能体群组工作机制与文献[21]中采用多智能体协同寻优的工作机制有较大区别。文献[21]偏重于算法设计,而文中更倾向于从物理层面将全网多无功调控动作分解为多个智能体各自的无功调控动作。

2.4 智能体成员状态改善奖励设计

结合电网无功电压运行状态,文中设定期望电压幅值为 1.0 p.u.。执行动作前后的电压幅值分别为 U_s 和 U_{s-} 。状态改善奖励设计如下:

(1) 若智能体执行动作后, U_{s-} 比 U_s 靠近期望电压,则智能体执行动作后的状态比动作前的状态有改善,奖励值为 1;

(2) 若智能体执行动作后, U_{s-} 比 U_s 远离期望电压,则智能体执行动作后的状态比动作前的状态有恶化,奖励值为 -1;

(3) 若智能体执行动作前后电压幅值距离期望电压幅值远近一致,则电网电压状态既无改善也未恶化,奖励值为 0。

以数学形式进行分类描述,奖励值 R 如式(5)所示。

$$R = \begin{cases} 1 & |U_{s-} - 1.0| < |U_s - 1.0| \\ 0 & |U_{s-} - 1.0| = |U_s - 1.0| \\ -1 & |U_{s-} - 1.0| > |U_s - 1.0| \end{cases} \quad (5)$$

由式(5)可见,文中设计的状态改善奖励机制完全基于 $t+1$ 时段电网状态与 t 时段状态相比的“ Δ ”量。且动作奖励值仅为“0”,“1”或“-1”,奖励方式简洁清晰,有助于智能体 Q 表快速收敛。

基于时序动态强化学习的电网无功电压调控方法优化流程见图 3。

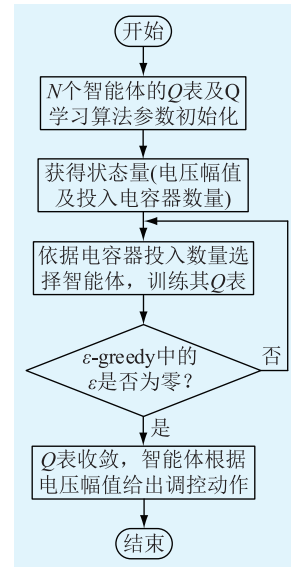


图 3 多智能体优化流程

Fig.3 Multi-agent optimization process

3 算例分析

3.1 算例描述

采用改进的 IEEE 10 机 39 节点系统验证方法可行性,改进系统见图 4。由离发电机较远的 8 节

点引出 40 节点作为负荷节点(图 4 中红圈),该节点设有 3 个电容器,容量均为 0.6 p.u.。电网中包括节点 40 在内共有 8 个节点有电容器。

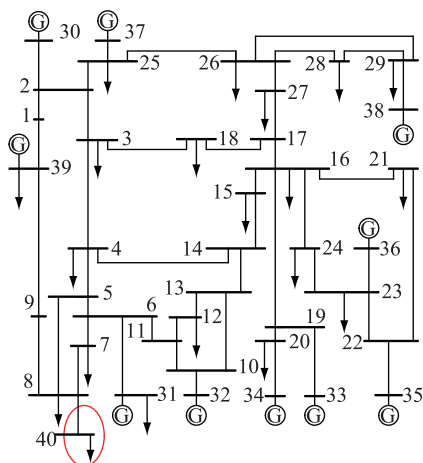


图 4 改进 IEEE 39 节点算例

Fig.4 Modified IEEE New England test system

3.2 单一智能体模式无功电压调控

采用全网只有一个智能体方式进行无功电压调控^[16]。每个变电站均可执行投切电容器动作(每次只投切 1 个电容),则该智能体动作集为 512 种,最终生成 Q 表见表 3。

表 3 单一智能体模式下的 Q 值

Table 3 Control strategies and Q values

状态名称	动作	Q 值
State-1	[节点 3 切电容,节点 5 切电容,节点 6 切电容,⋯,节点 40 切电容]	-1
	⋯	⋯
State-3	[节点 3 投电容,节点 5 投电容,节点 6 投电容,⋯,节点 40 投电容]	-1
	⋯	⋯
State-3	[节点 3 切电容,节点 5 切电容,节点 6 切电容,⋯,节点 40 切电容]	-1
	⋯	⋯
State-3	[节点 3 投电容,节点 5 投电容,节点 6 投电容,⋯,节点 40 投电容]	0.95
	⋯	⋯

通过以下 2 种工况对所得的单一智能体模式进行详细分析。

(1) 节点 40 电压幅值为 0.856 p.u.,该节点当前状态下无电容器投入。此时电压状态对应表 1 中电压幅值在 $(-\infty, 0.95]$ 的工况,电压极低。单一智能体模式动作集中的所有动作策略都不能让电压状态落入 0.95~1.05 p.u.。由于在奖励设计上将动作后的越限状态的回报值均设置成 -1.0,无法反映执行不同动作策略后状态之间的差异,因此该情况下智能体无法根据 Q 表给出最优动作(本例最优策略应为节点 40 投电容)。

(2) 节点 40 电压幅值为 0.963 p.u.,且该节点

已投入 3 个电容器。此时电压状态对应表 1 中电压幅值在 $(0.96, 0.97]$ 的工况。智能体根据表 3 中 Q 值结果选择投 1 个电容器。由于单一智能体模式的设计忽略了电网状态与电容器动作耦合关系,此时节点 40 已达投入电容器上限且无电容器可投,故智能体所做决策为无效决策。

以上分析可见,单一智能体模式无功电压调控方法在动作集和奖励函数设计上存在不足,在电网极端运行方式下无法给出合理动作。

3.3 智能体群组模式无功电压调控

按文中方法,各含电容器节点对应 1 个智能体子群组,全网共 8 个智能体子群组。每个子群组根据其所含电容器数量 m 又进一步细分为 $m+1$ 个智能体成员。节点 40 对应的子群组 $agent_{(40)}$ 所含智能体成员为 $\{agent_{(40),1}, agent_{(40),2}, agent_{(40),3}, agent_{(40),4}\}$ 。4 个智能体成员分别对变电站投入电容器的数量为 $\{0,1,2,3\}$ 。本例中因节点 40 距离其他节点较远,节点 40 电压仅由 $agent_{(40)}$ 进行调控。综合考虑动态无功电压调控的寻优空间和奖励值对当前状态的影响,式(3)中 ε 初值取为 0.2, ε 按照二次函数衰减,如图 5 所示。式(4)中 α 取值为 0.01, γ 取值为 0.9。前 15 000 次迭代, ε 衰减慢,增加了随机动作的几率,有利于避免局部最优解;后 5 000 次迭代, ε 衰减快,有利于加快 Q 表的收敛。经过 20 000 次迭代后, ε 衰减为 0,此时智能体不再随机选择动作, Q 表收敛。

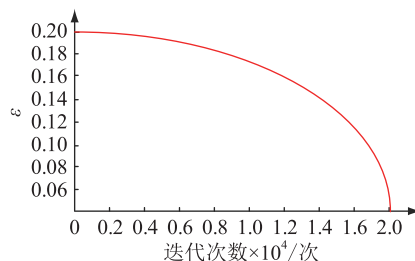


图 5 epsilon 衰减图

Fig.5 Epsilon decay graph

本算例中节点 40 接入负荷的有功功率在区间 $(0,4.5)$ 随机产生,无功功率在区间 $(-2,5)$ 随机产生。在电网无功电压时序变化状态下,按照式(4)的 Q 值更新策略得到表 4—表 7 的 4 个 Q 表。

智能体成员所感知的系统状态包含电压幅值与已投电容 2 个特征。各智能体成员动作仅为 2 或 3 种,智能体成员动作集受所感知状态影响而有所区别,且动作维度较单一智能体模式得以有效降低。受变电站无功补偿电容数量限制,当全部电容器投入且电压幅值位于区间 $(-\infty, 0.95]$ 时,智能体

表4 agent_{(40),1}对应Q表Table 4 The Q table of agent_{(40),1}

电压幅值/p.u.	动作		电压幅值/p.u.	动作	
	保持	投		保持	投
(-∞, 0.95]	0	1.946 5	(1.00, 1.01]	0	-0.972 3
(0.95, 0.96]	0	0.242 6	(1.01, 1.02]	0	-0.958 7
(0.96, 0.97]	0	0.214 3	(1.02, 1.03]	0	-0.961 5
(0.97, 0.98]	0	0.245 3	(1.03, 1.04]	0	-0.953 4
(0.98, 0.99]	0	-0.472 3	(1.04, 1.05]	0	-0.969 4
(0.99, 1.00]	0	-0.964 4	(1.05, +∞)	0	-1.000 0

表5 agent_{(40),2}对应Q表Table 5 The Q table of agent_{(40),2}

电压幅值/p.u.	动作		
	保持	投	切
(-∞, 0.95]	0	1.520 4	-0.245 3
(0.95, 0.96]	0	0.376 9	-0.077 3
(0.96, 0.97]	0	0.268 0	-0.019 9
(0.97, 0.98]	0	0.303 7	-0.029 7
(0.98, 0.99]	0	-0.198 8	-0.561 4
(0.99, 1.00]	0	-0.861 9	-0.864 7
(1.00, 1.01]	0	-0.836 2	-0.860 5
(1.01, 1.02]	0	-0.603 3	-0.376 1
(1.02, 1.03]	0	-0.067 9	0.413 0
(1.03, 1.04]	0	-0.067 9	0.401 0
(1.04, 1.05]	0	-0.077 3	0.401 0
(1.05, +∞)	0	-0.198 4	0.950 0

表6 agent_{(40),3}的Q表Table 6 The Q table of agent_{(40),3}

电压幅值/p.u.	动作		
	保持	投	切
(-∞, 0.95]	0	0.918 1	-0.165 5
(0.95, 0.96]	0	0.418 8	-0.010 0
(0.96, 0.97]	0	0.484 9	-0.039 4
(0.97, 0.98]	0	0.418 8	-0.049 0
(0.98, 0.99]	0	-0.508 2	-0.760 0
(0.99, 1.00]	0	-0.863 3	-0.834 5
(1.00, 1.01]	0	-0.856 3	-0.801 7
(1.01, 1.02]	0	-0.644 8	-0.423 4
(1.02, 1.03]	0	-0.077 3	0.350 9
(1.03, 1.04]	0	-0.049 0	0.337 7
(1.04, 1.05]	0	-0.010 0	0.255 6
(1.05, +∞)	0	-0.206 4	1.569 8

群组因当前全部电容器投入而选择 agent_{(40),4}, agent_{(40),4}最优动作为保持当前状态(因无其他电容可投);同理在电压幅值位于区间(1.05, +∞)当全部电容器已切时,智能体子群组 agent₍₄₀₎因全部电容器已切而选择 agent_{(40),1}, agent_{(40),1}的最优动作为保持当前状态(因无其他电容可切)。由此可见,当

表7 agent_{(40),4}对应Q表Table 7 The Q table of agent_{(40),4}

电压幅值/p.u.	动作		电压幅值/p.u.	动作	
	保持	投		保持	投
(-∞, 0.95]	0	-1.000 0	(1.00, 1.01]	0	-0.989 4
(0.95, 0.96]	0	-0.959 5	(1.01, 1.02]	0	-0.798 1
(0.96, 0.97]	0	-0.989 2	(1.02, 1.03]	0	0.230 3
(0.97, 0.98]	0	-0.980 0	(1.03, 1.04]	0	0.173 7
(0.98, 0.99]	0	-0.988 6	(1.04, 1.05]	0	0.214 7
(0.99, 1.00]	0	-0.994 9	(1.05, +∞)	0	1.965 2

智能体动作结果仍无法使得电网状态达标时,基于状态改善奖励设计的智能体仍能得到最优动作,体现了文中奖励设计机制的有效性。

图6为利用训练生成的节点40对应的智能体实现连续无功电压调控的效果。 U_{pre} 为未调控优化的电压幅值; U_{r1} 为按文献[16]设计的单一智能体模式下调控优化后的电压幅值; U_{r2} 为智能体群组模式下调控优化后的电压幅值。由仿真结果可见单一智能体模式在某些特殊工况下无法给出最优动作或给出的是无效动作。文中设计的智能体群组模式下调控优化步骤如下:连续动态无功电压调控任务开始的0时刻负荷中心节点已经接入1个电容器,子群组选择 agent_{(40),2},根据节点电压值对照 agent_{(40),2}的Q表执行动作(投入电容器),负荷中心节点电压变为0.99 p.u.;接着,由于节点40已投入2个电容器,群组选择 agent_{(40),3}。根据 agent_{(40),3}的Q表执行动作(保持),以此类推,每一步群组选择相应智能体并给出决策,最终实现电网全过程的动态无功电压调控。

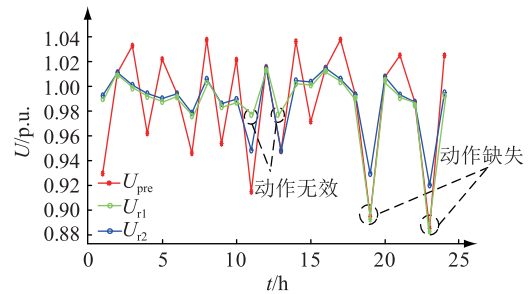


图6 某天24时刻的动态无功电压调控

Fig.6 Sequential reactive voltage regulation at 24 hours of a day

4 结语

文中针对无功电压调控具有多约束、非线性及非凸等特点,提出一种基于智能体群组强化学习的电网无功电压调控方法。通过分析现有研究存在的问题,将电容器投切纳入电网运行状态描述。设

计了一种强化学习智能体群组,以电容器投入数量作为智能体成员选择判据。群组中各智能体成员综合感知电压状态与电容器投入数量,各成员动作集维度较低,不同成员动作集也存在区别。所提出状态改善奖励机制可在极端运行工况下有效筛选出最优动作。文中智能体群组设计可克服单一智能体模式诸多技术问题。

智能体获取电网的运行状态信息基于对电网运行态势的充分感知,而电网状态描述一直是人工智能领域的难题。基于深度强化学习算法的智能体在感知能力上较传统强化学习智能体更具优势,未来将做进一步研究。

参考文献:

[1] 杨恩惠. 对当前我国电网无功电压的调查分析[J]. 电网技术,1992(2):22-27.
 YANG Enhui. The investigation and analysis of reactive power voltage of power grid in China[J]. Power System Technology, 1992(2):22-27.

[2] 耿少博,顾乔根,常风然,等. 交直流混合配电网分布式无功电压互动控制策略[J]. 电力工程技术,2019,38(4):26-33.
 GENG Shaobo, GU Qiaogen, CHANG Fengran, et al. Distributed reactive voltage and voltage interactive control strategy for AC/DC hybrid distribution network [J]. Electric Power Engineering Technology, 2019,38(4):26-33.

[3] 张勇军,俞悦,任震,等. 实时环境下动态无功优化建模研究[J]. 电网技术,2004,28(12):12-15.
 ZHANG Yongjun, YU Yue, REN Zhen, et al. Research on dynamic modeling for reactive power optimization under real-time circumstance [J]. Power System Technology, 2004, 28(12):12-15.

[4] 邓佑满,张伯明,相年德. 配电网电容器实时优化投切的逐次线性整数规划法[J]. 中国电机工程学报,1995,15(6):375-383.
 DENG Youman, ZHANG Boming, XIANG Niande. A successive linear integer programming methodology for capacitor switching on distribution system [J]. Proceedings of the CSEE, 1995, 15(6):375-383.

[5] 楼伯良,杨少杰,叶琳,等. 大容量无功补偿装置在浙江受端电网的落点优选方法[J]. 广东电力,2019,32(10):96-101.
 LOU Boliang, YANG Shaojie, YE Lin, et al. Optimization method for large capacity reactive power compensation device in Zhejiang receiving end power grid [J]. Guangdong Electric Power, 2019,32(10):96-101.

[6] 肖繁,王涛,高扬,等. 基于特高压交直流混联电网的调相机无功补偿及快速响应机制研究[J]. 电力系统保护与控制,2019,47(17):93-100.
 XIAO Fan, WANG Tao, GAO Yang, et al. Research on reactive power compensation and fast response mechanism of synchronous condenser based on UHVAC/DC hybrid grid [J]. Power System Protection and Control, 2019,47(17):93-100.

[7] 任晓娟,邓佑满,赵长城,等. 高中压配电网动态无功优化算法的研究[J]. 中国电机工程学报,2003,23(1):32-37.
 REN Xiaojuan, DENG Youman, ZHAO Changcheng, et al. Study on the algorithm for dynamic reactive power optimization of distribution systems [J]. Proceedings of the CSEE, 2003, 23(1):32-37.

[8] YAMIN H Y. Dynamic optimal power flow using interior point method and benders decomposition considering active and reactive constraints [J]. Electric Power Components and Systems, 2006,34(12):1377-1393.

[9] 赖永生,刘明波. 电力系统动态无功优化问题的快速解耦算法[J]. 中国电机工程学报,2008,28(7):32-39.
 LAI Yongsheng, LIU Mingbo. Fast decomposition algorithm for solution of dynamic reactive power optimization problem in power systems [J]. Proceedings of the CSEE, 2008, 28(7):32-39.

[10] 盛戈皞,涂光瑜,罗毅. 人工智能技术在电力系统无功电压控制中的应用[J]. 电网技术,2002,26(6):22-27.
 SHENG Yihao, TU Guangyu, LUO Yi. Application of artificial intelligence techniques in reactive power/voltage control of power system [J]. Power System Technology, 2002, 26(6):22-27.

[11] 刘方,颜伟. 基于遗传算法和内点法的无功优化混合策略[J]. 中国电机工程学报,2005,25(15):67-72.
 LIU Fang, YAN Wei. A hybrid strategy based on GA and IPM for optimal reactive power flow [J]. Proceedings of the CSEE, 2005, 25(15):67-72.

[12] 鲁裕婷,赵天乐,都洪基,等. 基于改进粒子群算法的含DG配电网无功优化[J]. 电力工程技术,2018,37(6):69-74.
 LU Yuting, ZHAO Tianle, DU Hongji, et al. Reactive power optimization of distribution network with distributed generation based on improved particle swarm optimization algorithm [J]. Electric Power Engineering Technology, 2018,37(6):69-74.

[13] 周念成,廖建权,王强钢,等. 深度学习在智能电网中的应用现状分析与展望[J]. 电力系统自动化,2019,43(4):180-197.
 ZHOU Niancheng, LIAO Jianquan, WANG Qianggang, et al. Analysis and prospec of deep learning application in smart grid [J]. Automation of Electric Power System, 2019, 43(4):180-197.

[14] 吴双,胡伟,张林,等. 基于AI技术的电网关键稳定特征智能选择方法[J]. 中国电机工程学报,2019,39(1):14-21,316.
 WU Shuang, HU Wei, ZHANG Lin, et al. An intelligent key feature selection method of power grid based on artificial intelligence technology [J]. Proceedings of the CSEE, 2019, 39(1):14-21,316.

[15] WANG Songyan, FAN Shixiong, CHEN Jianwen, et al. Deep-learning based fault diagnosis using computer-visualised power flow [J]. IET Generation, Transmission & Distribution, 2018, 12(17):3985-3992.

[16] 刁浩然,杨明,陈芳,等. 基于强化学习理论的地区电网无

- 功电压优化控制方法[J]. 电工技术学报, 2015, 30(12): 408-414.
- DIAO Haoran, YANG Ming, CHEN Fang, et al. Reactive power and voltage optimization control approach of the regional power grid based on reinforcement learning theory[J]. Transactions of China Electrotechnical Society, 2015, 30(12): 408-414.
- [17] ZHANG Xiaoshun, YU Tao, YANG Bo, et al. Accelerating bio-inspired optimizer with transfer reinforcement learning for reactive power optimization[J]. Knowledge-based Systems, 2017, 116(1): 26-38.
- [18] 郭乐欣. 基于多主体强化学习的碳能复合流无功优化算法研究[D]. 广州:华南理工大学, 2016.
- GUO Lexin. Research of carbon-energy combined-flow reactive power optimization based on multi-agent reinforcement learning algorithm[D]. Guangzhou: South China University of Technology, 2016.
- [19] 孙国忠. 基于渐进学习理论的配电网无功电压控制方法研究[D]. 济南:山东大学, 2014.
- SUN Guozhong. Research on the approach of the reactive power and voltage control of distribution network based on incremental learning theory[D]. Jinan: Shandong University, 2014.
- [20] ZHANG Xiaoshun, YU Tao, YANG Bo, et al. Approximate ideal multi-objective solution $Q(\lambda)$ learning for optimal carbon-energy combined-flow in multi-energy power systems [J]. Energy Conversion and Management, 2015, 9(49): 543-556.
- [21] 陈艺璇, 张孝顺, 郭乐欣, 等. 基于多智能体迁移强化学习算法的电力系统最优碳-能复合流求解[J]. 高电压技术, 2019, 45(3): 863-872.
- CHEN Yixuan, ZHANG Xiaoshun, GUO Lexin, et al. Optimal carbon-energy combined flow in power system based on multi-agent transfer reinforcement learning[J]. High Voltage Engineering, 2019, 45(3): 863-872.
- [22] 张孝顺. 电力系统的迁移强化学习优化算法研究[D]. 广州:华南理工大学, 2017.
- ZHANG Xiaoshun. Transfer reinforcement learning for power system optimization[D]. Guangzhou: South China University of Technology, 2017.

作者简介:



范士雄

范士雄(1984),男,博士,高级工程师,从事电网调控和人工智能技术相关工作(E-mail: fanshixiong@epri.sgcc.com.cn);

刘幸蔚(1991),女,硕士,工程师,从事电网调控和人工智能技术相关工作;

魏智慧(1995),男,硕士在读,研究方向为无功电压智能调控技术。

Reactive voltage regulation method based on agents group using reinforcement learning

FAN Shixiong¹, LIU Xingwei¹, WEI Zhihui², LIU Ruiye², WANG Songyan², YU Jilai²

(1. China Electric Power Research Institute Co., Ltd., Beijing 100192, China;

2. School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Reactive voltage regulation is an important control measure to ensure voltage stability in the power grid area. The existing single agent design method based on reinforcement learning has many problems, such as high coupling between state and action, various combinations of reactive power compensation devices, and unreasonable reward design based on target deviation model. Aiming at these problems, a method for describing the integrated operation state of the grid considering the node voltage amplitude and capacitor switching condition is proposed. The reinforcement learning agent group architecture for grid reactive voltage regulation is designed. The group determines the members of the agent according to the current comprehensive operating state of the grid, and gives corresponding reactive power regulation actions. Each agent member uses the improvement degree of the grid state in the adjacent time period as a reward mechanism. The example shows that the method can be applied to the grid reactive voltage regulation environment. Compared with the single agent design method, the number of action sets can be effectively reduced, and various reactive voltage regulation conditions can be better dealt with.

Keywords: reactive power and voltage; reinforcement learning; artificial intelligence; agent group; capacitor switch

(编辑 方晶)