

基于深度强化学习的 Boost 变换器控制策略

戴宇轩, 崔承刚*

(上海电力大学 自动化学院, 上海 200090)

摘要: 针对 Boost 变换器母线电压稳定控制存在模型不确定性和非线性问题, 提出了一种基于无模型深度强化学习的智能控制策略。结合强化学习 DDQN(double DQN) 算法与 DDPG 算法设计了 Boost 变换器控制器, 包括了状态、动作空间、奖励函数以及神经网络的设计以提高控制器动态性能; 基于 ModelicaGym 库开发工具包 reinforment learning modelica(RLM) 实现了 Boost 变换器模型与强化学习智能体的联合仿真。通过与双环 PI 控制器的对比仿真表明: 强化学习控制器在三种工况下的母线电压稳定控制结果具有更好的动态性能。

关键词: Boost 变换器; 深度强化学习; DDQN 算法; DDPG 算法; 协同仿真

中图分类号: TP391.9

文献标志码: A

文章编号: 1004-731X(2023)05-1109-11

DOI: 10.16182/j.issn1004731x.joss.22-0101

引用格式: 戴宇轩, 崔承刚. 基于深度强化学习的 Boost 变换器控制策略[J]. 系统仿真学报, 2023, 35(5): 1109-1119.

Reference format: Dai Yuxuan, Cui Chenggang. Deep Reinforcement Learning-Based Control Strategy for Boost Converter[J]. Journal of System Simulation, 2023, 35(5): 1109-1119.

Deep Reinforcement Learning-Based Control Strategy for Boost Converter

Dai Yuxuan, Cui Chenggang*

(College of Automation Engineering, Shanghai University of Electric Power, Shanghai 200090, China)

Abstract: In view of the problems of model uncertainty and nonlinearity in bus voltage stability control of Boost converter, an intelligent control strategy based on model-free deep reinforcement learning(RL) is proposed. RL double DQN(DDQN) algorithm and deep deterministic policy gradient(DDPG) algorithm are used, and the Boost converter controller is designed. The state, action space, reward function, and neural network are also designed to improve the dynamic performance of the controller. The joint simulation of the Boost converter model and RL agent is realized by RL modelica(RLM), a toolkit developed based on ModelicaGym. The proposed controller is compared with the double-loop PI controller, and the simulation shows that the bus voltage stability control based on the RL controller has better dynamic performance under three working conditions.

Keywords: Boost converter; deep reinforcement learning; DDQN (double DQN) algorithm; DDPG (deep deterministic policy gradient) algorithm; joint simulation

0 引言

DC-DC 变换器是最常见的电力电子变换器。为实现高功率转换效率以及保持供电稳定, 各国

学者对其控制策略进行了大量研究工作。例如反步控制^[1-2], 模型预测控制^[3-4]和滑模控制^[5-6]等。当 DC-DC 变换器为恒功率负载(constant power load,

收稿日期: 2022-02-10

修回日期: 2022-03-18

基金项目: 国家自然科学基金青年科学基金(5160711); 上海市 2021 年度“科技创新行动计划”科技支撑碳达峰碳中和专项(21DZ1207302); 上海市科学技术委员会科研计划(19DZ1205700)

第一作者: 戴宇轩(1997-), 男, 硕士生, 研究方向为人工智能在直流微电网中的应用。E-mail: 475817594@qq.com

通讯作者: 崔承刚(1981-), 男, 副教授, 博士, 研究方向为分布式能源系统预测与优化调度。E-mail: cgcui@shiep.edu.cn

CPL)供电时, 由于CPL的负阻抗特性给电路带来的不稳定特性和非线性, 传统控制方法难以平衡稳定时间和超调^[7], 基于模型的方案所需的计算时间较长, 无法实时解决。

对于诸如Boost变换器这类具有的非线性, 强耦合特性的电力电子系统, 面对工况变化时, 确定性方法的控制性能会下降, 因此较难设计控制器来达到理想的动态特性。在前人的研究中, 智能控制器包括模糊PID控制^[8-9], 模糊神经网络控制^[10], 基于脑情感学习的智能控制器^[11]等方法在适应系统的可变条件方面具有明显的优势。因此, 将智能算法作为控制策略近年来也引起了广泛的关注。

强化学习(reinforcement learning, RL)作为智能算法一个新热点越来越受到重视。RL作为一种新兴的优化控制策略算法, 可以解决各类复杂的任务, 例如四旋翼姿态控制^[12], 机器人运动控制^[13]以及无人机集群的控制^[14]。在电力系统控制方面, 强化学习被用于综合能源管理^[15]与潮流优化控制^[16]。在DC-DC变换器控制方面, 文献[17]提出了一种将RL算法Actor-Critic并入超局部模型(ULM)的控制方案。文献[18]使用DDPG(deep deterministic policy gradient)算法为滑模观测器提供补偿器信号从而实现DC-DC Buck-Boost变换器控制。然而, 以上方法都是通过与传统控制方法结合实现DC-DC变换器控制。故仍然依赖于DC-DC变换器数学模型。而具有更强的自适应性和泛化性的无模型RL算法Boost变换器控制设计的相关文献还鲜有报导。

针对Boost变换器控制的时变性和非线性特性, 本文提出了一种基于深度强化学习的无模型控制策略。首先设计了基于强化学习的Boost变换器控制器, 选取电路中电压信息为状态, 开关器件的占空比为动作, 以降低母线电压误差为目标设计奖励函数。其次, 在ModelicaGym^[19]的基础上开发了RLM(reinforcement learning modelica)工具包, 用于连接Dymola平台中Modelica模型与

Python中OpenAI Gym库进行协同仿真。最后将两种深度强化学习控制器与双环PI控制器在微源电压变化、阻性负载以及恒功率负载变化三种工况下进行仿真比较, 验证强化学习算法在快速跟踪母线电压时动态性能的优越性。

1 基于Modelica的Boost变换器建模

1.1 Boost变换器拓扑及原理

当负载为阻性负载时, Boost变换器在电流连续模式下的动态方程为

$$\begin{cases} L \frac{di_L}{dt} = E - (1-d)V_o \\ C \frac{dV_o}{dt} = (1-d)i_L - \frac{V_o}{R} \end{cases} \quad (1)$$

式中: i_L 和 $V_o \in \mathbb{R} > 0$ 分别表示电感电流和电容电压的平均值, $E \in \mathbb{R} > 0$ 表示微源电压, C 和 L 分别为升压变换器的电容和电感, 控制输入由占空比 $d \in \{0, 1\}$ 表示。

当带载为CPL时, 有 $i_L = \frac{P}{V_o}$, 动态方程变为

$$\begin{cases} L \frac{di_L}{dt} = E - (1-d)V_o \\ C \frac{dV_o}{dt} = (1-d)i_L - \frac{P}{V_o} \end{cases} \quad (2)$$

为保证负载稳定运行, 电源向CPL提供的功率需保持恒定。可见CPL的瞬时阻抗是正值, 即 $\frac{V_o}{i_L} > 0$ 。但如在运行点 (i_L, V_o) 对电流求导可得

$$\frac{dV_o}{di_L} = \frac{d\left(\frac{P}{i_L}\right)}{di_L} = -\frac{P}{i_L^2} = -\frac{V_o}{i_L} \quad (3)$$

1.2 基于Dymola仿真平台建立的仿真模型

Dymola 是法国 Dassault Systems 公司的多学科系统仿真平台, 广泛应用于国内外汽车、工业、能源等行业的系统功能验证及优化。Dymola 支持 FMI 标准接口协议, 具有多领域建模与非因果建模等特性。本文在Dymola平台搭建基于Modelica

语言的 Boost 变换器电路模型, 图 1 为 Dymola 中 Boost 变换器模型图。

2 基于深度强化学习的控制器设计

2.1 问题描述

针对在不同干扰下母线电压快速跟踪控制问

题, 本文提出了两种基于强化学习(RL)的 Boost 变换器控制方案。图 2 展示了系统整体控制框图, RL 控制器根据电压信息计算出动作对应的占空比输出, 并生成 PWM 控制开关器件的通断。RL 控制器与环境交互, 执行每个动作后, 下一时刻继续采样挑选动作, 从而实现母线电压快速跟踪控制。

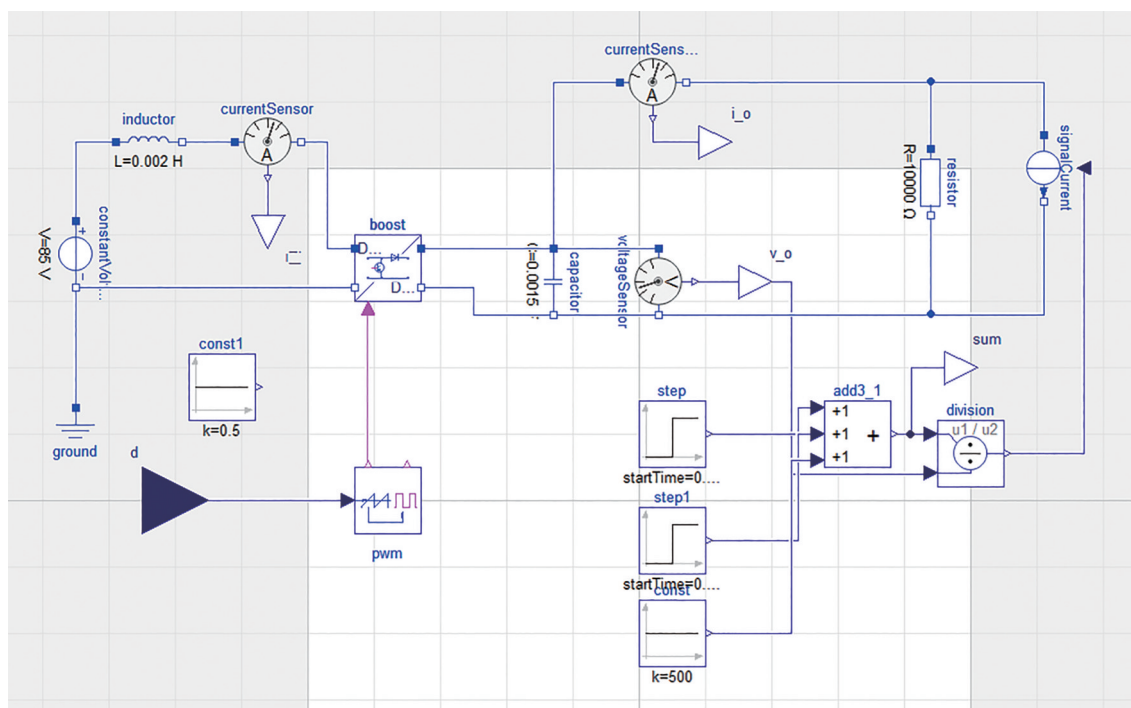


图 1 基于 Dymola 平台的 Boost 变换器
Fig. 1 Boost converter based on Dymola platform

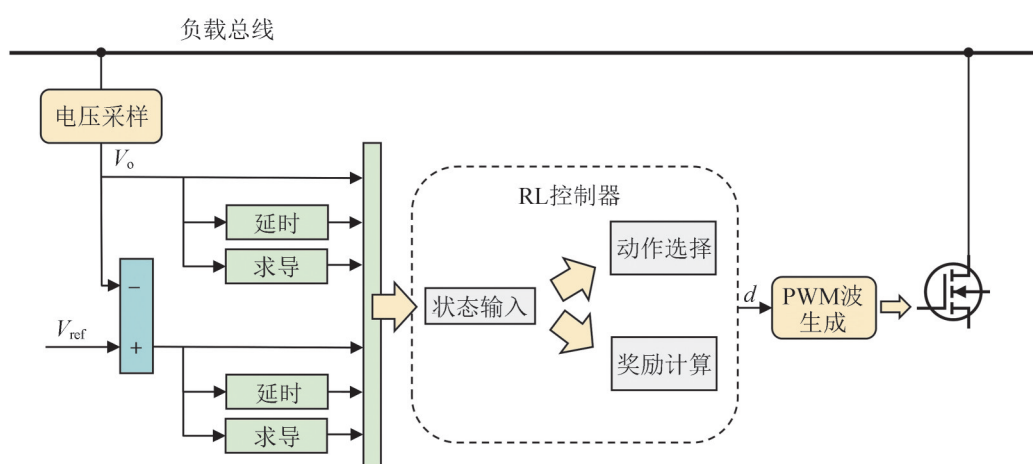


图 2 强化学习控制器整体控制框图
Fig. 2 Overall control block diagram of RL controller

2.2 Double DQN控制器设计

本文中使用的第一种算法为DQN的改进算法DDQN(double DQN)，流程图如图3所示。

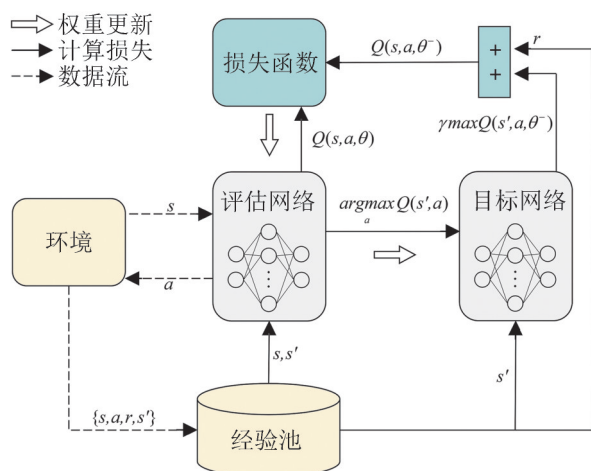


图3 DDQN流程图
Fig. 3 Flow chart of DDQN

首先，引入神经网络对当前状态-动作对的 Q 值 $Q(s, a|\theta)$ 进行估计，其中 θ 表示神经网络中的参数。输入为离散状态信息，输出为状态动作函数 y 的估计值 \hat{y} ，选择 \hat{y} 最大的动作输出。然后，与环境交互执行每个动作后，将状态转移向量 $\{s, a, r, s'\}$ 存入经验池。最后，从经验池中随机进行批采样数据来更新评估网络参数，同时对目标网络进行软更新。本文中的第一种算法相关设计如下：

(1) 状态空间

作为神经网络的输入，状态空间的选择需要包含足够的信息来充分描述系统环境。为了消除状态之间的量纲影响。本文对状态做了归一化处理，使得各状态分量尽量处于同一数量级内：

$$S = \left\{ \frac{V_o}{170}, \frac{V_{o_delay}}{170}, \frac{dV_o}{dt}, e, e_{delay}, \frac{de}{dt} \right\} \quad (4)$$

其中 V_o 为输出端的负载电压， V_{o_delay} 为上一时刻的负载电压， $e = V_{ref} - V_o$ 为误差电压， $\frac{dV_o}{dt}$ ， $\frac{de}{dt}$ 分别为负载电压和误差电压的导数。

(2) 动作空间

第一种算法中动作输出为占空比 $d \in (0, 1)$ ，

DDQN动作空间为占空比的离散集合，具体数值见各工况介绍：

$$A = \{d_1, d_2, \dots, d_{n_a}\} \quad (5)$$

(3) 奖励函数

考虑到本文控制目标是在最短的时间内使输出电压误差最小。因此，第一种算法采用离散奖励函数，将每个时间步中的输出电压偏差绝对值乘上一个负的系数作为惩罚，奖励根据误差大小分为两段并配以不同系数：

$$\begin{aligned} e &= |V_{ref} - V_o| \\ \text{penalty} &= -10e \\ \text{reward} &= \begin{cases} 1, & e < 1 \\ 10, & e < 0.1 \end{cases} \\ R &= \text{penalty} + \text{reward} \end{aligned} \quad (6)$$

(4) 神经网络设计

DDQN使用神经网络逼近状态-动作函数，在状态 s 下选择最大 $Q(s, a)$ 对应的动作 a 。本文控制器中的神经网络由一个输入层、一个输出层以及双隐层组成，如图4所示。其中输入层的输入节点数为状态的数量 $n_s = 6$ ，输出层的输出节点数为离散动作个数 n_a (视工况而定)，双隐层神经元个数皆为 $n_h = 64$ ，除输出层外，激活函数皆选用ReLU修正线性单元。

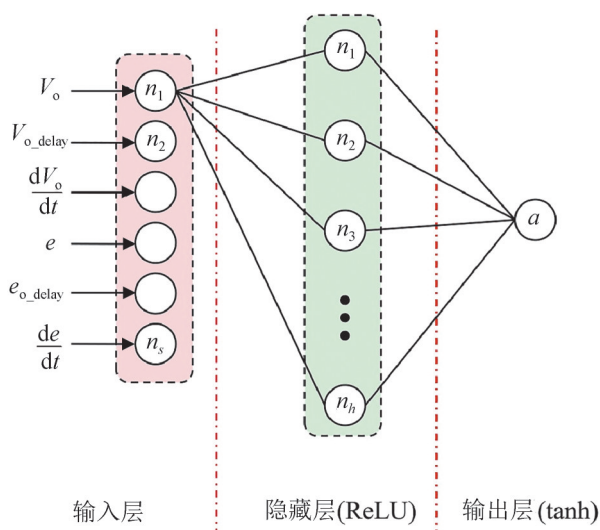


图4 DDQN神经网络设计
Fig. 4 DDQN Neural network design

2.3 DDPG 控制器设计

本课题用到的第二大算法即为 DDPG, 流程图如图 5 所示。智能体与环境交互时, 根据当前环境给出的状态 s 作为估计动作网络的输入, 网络的输出加上一个噪声形成动作 a 。执行动作后环境跳转到下一状态 s' 并返回一个即时奖励 r 。交互产生的经验不会直接用于更新, 而是存入经验池中, 单步执行完后抽取小批次的经验对网络参数进行更新。本文中 DDPG 算法相关设计如下:

(1) 状态空间

其状态空间设计与 DDQN 一致, 同样需要进行归一化处理:

$$S = \left\{ \frac{V_o}{170}, \frac{V_{o_delay}}{170}, \frac{dV_o}{dt}, e, e_{delay}, \frac{de}{dt} \right\} \quad (7)$$

(2) 动作空间

第二种算法中动作输出为放大倍数 N , 取值范围由基准值 $bias$ 与动作值 $action$ 决定:

$$N \in [bias - action, bias + action] \quad (8)$$

再通过放大倍数 N 计算占空比 d , 具体数值见

各工况介绍:

$$d = \frac{1}{1-N} \quad (9)$$

(3) 奖励函数

DDPG 采用连续奖励函数, 将每个时间步中的输出电压偏差绝对值乘上一个负的系数作为惩罚。奖励随着误差减小而增大, 单步奖励上限为 5。

$$\begin{aligned} e &= |V_{ref} - V_o| \\ penalty &= -10e \\ reward &= \frac{0.5}{0.05 + e} \\ R &= penalty + reward \end{aligned} \quad (10)$$

(4) 神经网络设计

DDPG 算法属于强化学习中的 Actor-Critic 架构, 故使用的神经网络包括两部分。Actor 动作网络输入为状态 s , 输出为动作 a 。中间隐藏层神经元数 $n_h = 128$, 如图 6 左半部分所示。其中隐藏层的激活函数为 ReLU, 为了使输出动作以 0 为中心对称, 输出层选用的激活函数为 tanh。

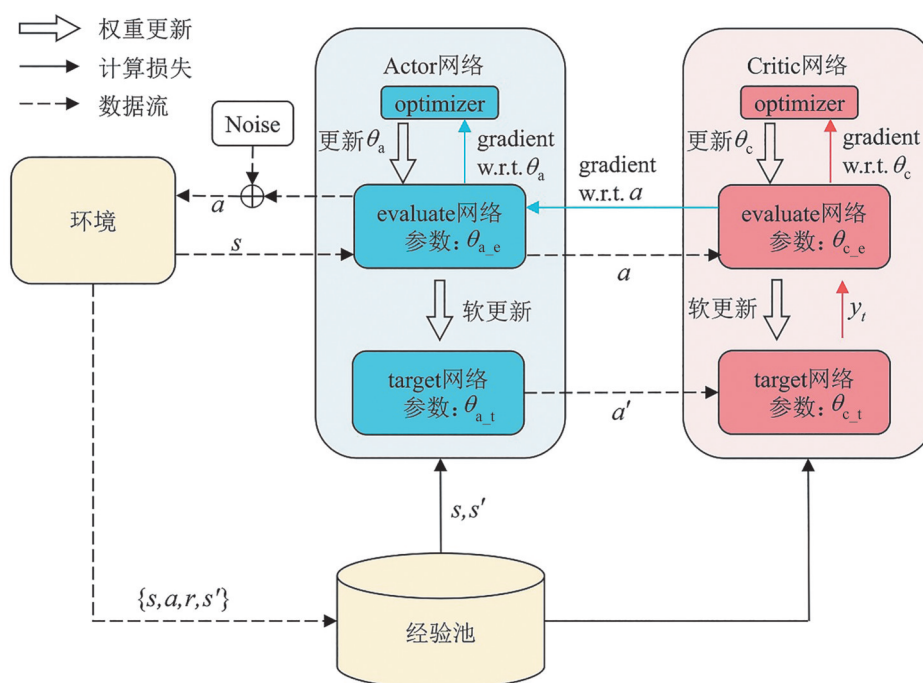


图 5 DDPG 流程图

Fig. 5 Flow chart of DDPG

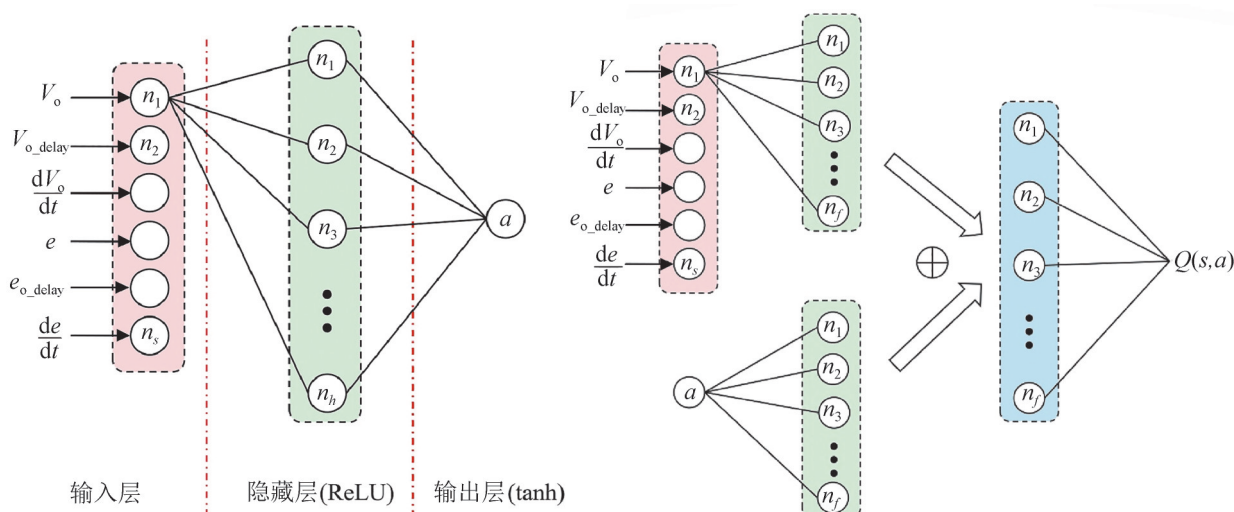


图 6 DDPG Actor-Critic 网络设计
Fig. 6 DDPG Actor-Critic network design

Critic 估计网络也与 DDQN 中的不同，如图 6 右半部分所示。其包含两条路径，首先是对状态抽取特征，特征数 $n_f=128$ 。再抽取动作特征，特征数 $n_f=128$ 。将两条路径抽取的特征叠加，再通过一层全连接计算当前状态-动作对的 Q 值 $Q(s, a)$ ，Critic 网络中使用的激活函数皆为 ReLU。

3 Reinforcement Learning Modelica (RLM)工具包

本文基于 Python 语言开发了 Reinforcement Learning Modelica(RLM)工具包。该工具包旨在采用强化学习 (RL) 来解决 Modelica 模型中的优化和控制任务。本工具包根据功能模型接口 (FMI2.0) 标准将基于 Modelica 语言的模型打包封装成 FMU 模块，再使用 Python 加载模型进行协同仿真。整体平台如图 7 所示。RLM 工具包具有以下优势：

- (1) 环境模块的编写遵循 OpenAI Gym 规范，可以适用于各种常见 RL 算法库。
- (2) 可以为不熟悉 Modelica 或 FMU 的用户提供用于解决现实世界问题的 RL 应用的环境。
- (3) 将自定义奖励函数、状态动作设计的功能集成到统一环境程序中，开箱即用。



图 7 RLM 工具包功能示意图
Fig. 7 Schematic diagram of RLM toolkit function

4 仿真实验与分析

4.1 系统仿真参数

为了验证强化学习控制器的性能，本文在 Dymola/Python 平台上搭建了 Boost 变换器仿真模型。系统具体参数如表 1 所示，表 2 给出了强化学习算法训练时使用的超参数。

表 1 电路参数
Table 1 Circuit parameters

参数	数值
电感 L/mH	2
电容 $C/\mu\text{F}$	1500
输入电压 E/V	85
母线电压 V_{ref}/V	170

4.2 仿真工况介绍

为了验证基于强化学习控制器的动态性能优越性，本文将提出的两种 RL 控制器与传统双 PI 算

法进行了控制效果的比较。共在微源电压变化、阻性负载变化和恒功率负载变化三种工况下进行对比。相关的参数均已根据仿真需求与人工经验进行调参优化处理, 各工况下的双 PI 参数见表 3。

表 2 RL 训练超参数
Table 2 RL training hyperparameters

参数	数值
学习率 α	0.001
DDQN 折扣率 γ_1	0.9
DDPG 折扣率 γ_2	0.98
批采样大小 b	256
经验池大小 B	2×10^5

表 3 三种工况下 PI 控制器参数设置
Table 3 PI controller parameter settings under three working conditions

工况	k_{vp}	k_{vi}	K_{cp}	K_{ci}
微源电压变化	100	30	0.005	2
阻性负载变化	30	500	0.01	10
恒功率负载变化	1	300	0.02	100

(1) 微源电压变化

工况一研究了微源电压波动的情况下 RL 控制器对于母线电压变化的追踪能力。在整个实时仿真中, $t \in [0, 0.08]s$ 的微源电压 $E=85V$, $t \in [0.08, 0.16]s$ 的微源电压 $E=75V$, $t \in [0.16, 0.24]s$ 的微源电压 $E=95V$, 负载为固定的阻性负载 $R=1000\Omega$ 。

该工况下 DDQN 动作空间为

$$A = \{0.23, 0.24, \dots, 0.47\} \quad (11)$$

DDPG 动作空间为

$$N \in [1.6 - 0.3, 1.6 + 0.3] \quad (12)$$

(2) 阻性负载变化

工况二研究了阻性负载接入断开的情况下 RL 控制器对于母线电压变化的追踪能力。在整个实时仿真中, $t \in [0, 0.08]s$ 的阻性负载为 1000Ω , $t \in [0.08, 0.16]s$ 的阻性负载为 500Ω , $t \in [0.16, 0.24]s$ 的阻性负载为 1000Ω 。

该工况下 DDQN 动作空间为

$$A = \{0.22, 0.23, \dots, 0.50\} \quad (13)$$

DDPG 动作空间为

$$N \in [1.6 - 0.3, 1.6 + 0.3] \quad (14)$$

(3) 恒功率负载变化

工况三研究了恒功率负载接入断开的情况下 RL 控制器对于母线电压变化的追踪能力。CPL 带有负阻抗特性, 在其功率改变时会对升压变换器造成不稳定和非线性的影响。本文使用受控电流源并联大电阻模拟 CPL 效果, 在整个实时仿真中, $t \in [0, 0.08]s$ 的恒功率负载功率 $P_{CPL} = 500W$, $t \in [0.08, 0.16]s$ 的恒功率负载功率 $P_{CPL} = 300W$, $t \in [0.16, 0.24]s$ 的恒功率负载功率 $P_{CPL} = 700W$ 。

该工况下 DDQN 动作空间为

$$A = \{0.44, 0.45, \dots, 0.55\} \quad (15)$$

DDPG 动作空间为

$$N \in [2 - 0.4, 2 + 0.4] \quad (16)$$

4.3 仿真分析

图 8 为工况一中模拟微源电压变化与工况二中模拟阻性负载变化的仿真结果。其中红色线代表基准母线电压, 蓝色线为双闭环 PI 控制器电压波形, 橙色线与绿色线分别为 DDQN 控制器与 DDPG 控制器电压波形。从图 8(a)可以看出, 面对三段微源电压变化时, 在传统双环 PI 控制下, 母线电压波动分别约为 $(0.30, 0.07, 0.12)V$, 稳定时间分别约为 $(30, 15, 20)ms$ 。从图 8(b)可以看出, 面对三段阻性负载变化时, 在传统双环 PI 控制下, 母线电压波动分别约为 $(0.31, 0.18, 0.15)V$, 稳定时间分别约为 $(35, 20, 30)ms$ 。

可以看到, 在系统面临的扰动非线性较弱的情况下, RL 控制器能更快实现母线电压追踪。与 PI 控制效果相比, 其超调与稳定时间都接近于 0。由于 DDPG 算法动作的连续性, 使得其动态性能与稳态性能都更胜一筹。

图 9 描述了工况三中模拟恒功率负载变化三种控制器的控制情况。其中红色线代表基准母线

电压, 蓝色线为双闭环 PI 控制器电压波形, 橙色线
线与绿色线分别为 DDQN 控制器与 DDPG 控制器
电压波形。从电压波动角度出发, 在 $t=0$ s、 $t=$
 0.08 s、 $t=0.16$ s 时, 由于负载变化, 母线电压发
生波动。三种控制器的电压波动分别为(4.7, 1.4,
2.8) V、(1.4, 0.4, 1.2) V、(0.8, 0.3, 1.0) V。

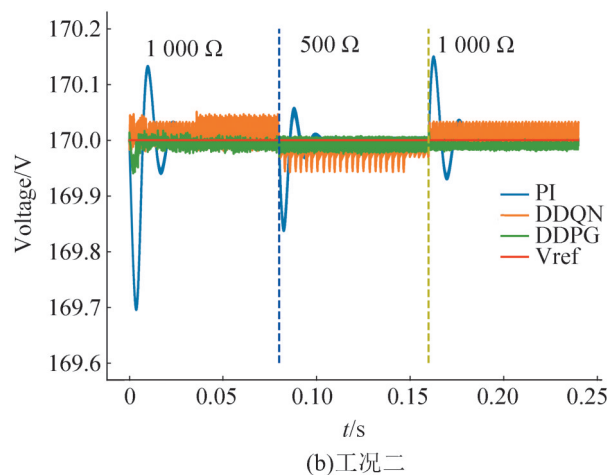
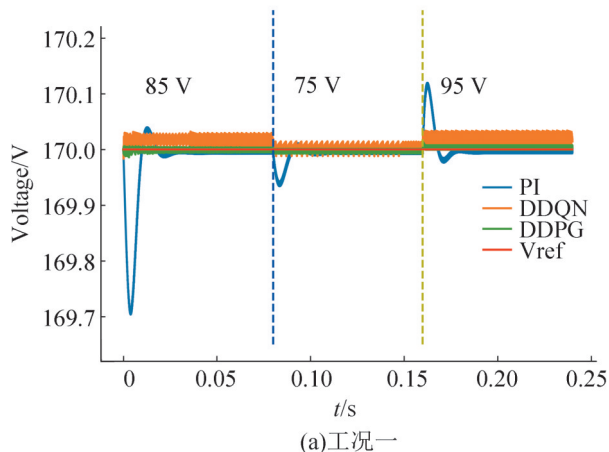


图 8 电压波形对比
Fig. 8 Comparison of voltage waveforms

从稳定时间角度出发, 三种控制器的稳定时
间分别为(30, 23, 23) ms、(10, 5, 5) ms、(5, 3, 4) ms,
如图 10 所示。

从电流波动角度出发, 三种控制器在工况三
下对电感电流的控制效果如图 11 所示, 可以看到
右边两种 RL 控制器控制下的电感电流能更快的恢
复稳定。

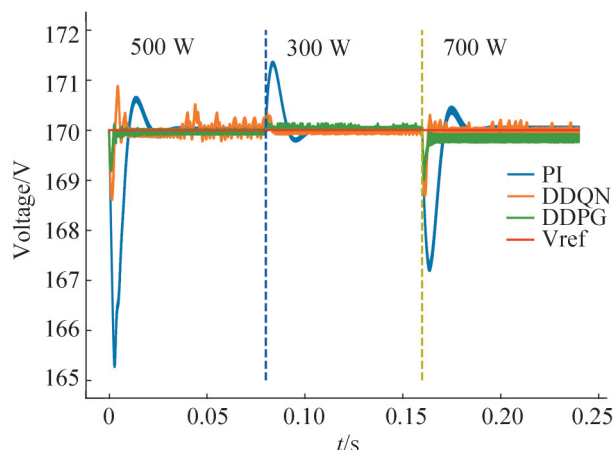


图 9 工况三下电压波形对比
Fig. 9 Comparison of voltage waveforms under working
condition three

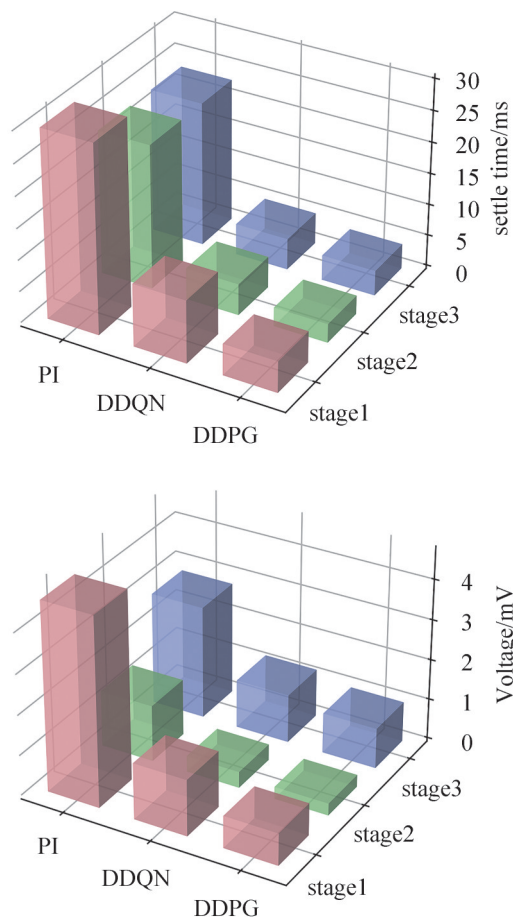


图 10 工况三下动态性能对比
Fig. 10 Dynamic performance comparison under working
condition three

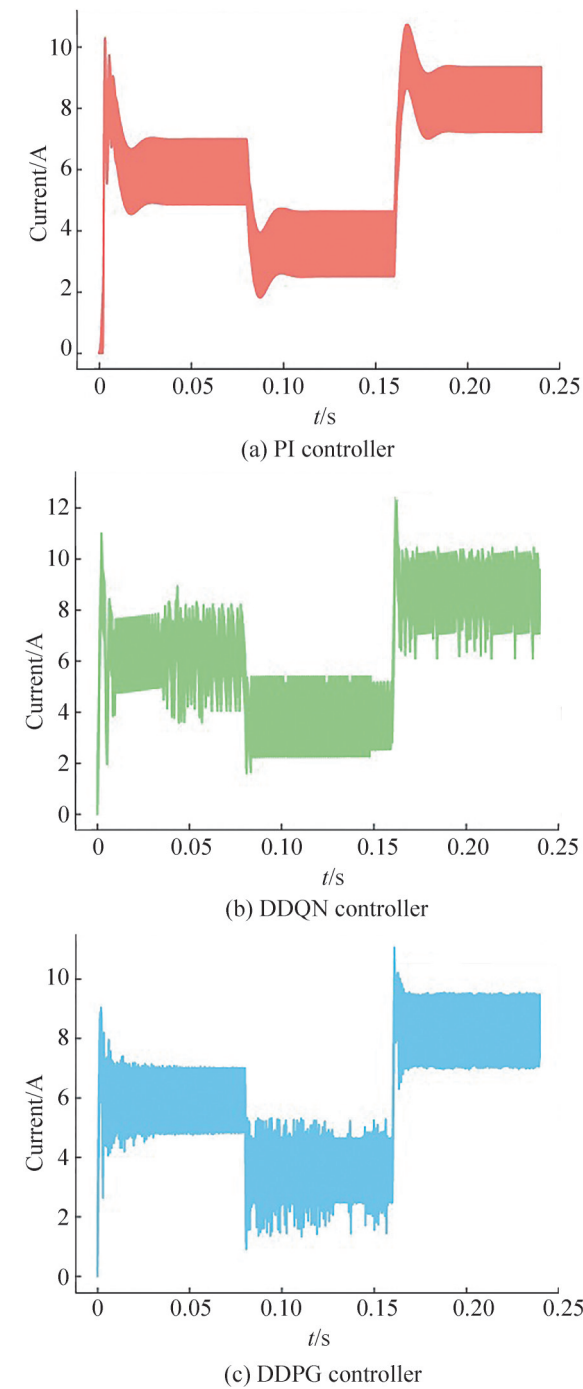


图 11 工况三下电感电流对比
Fig. 11 Comparison of inductive currents under working condition three

为了对两种控制器的整体误差进行定量分析, 图 12 使用了三种误差衡量标准量化电压波动程度。包括积分误差、平均绝对误差和均方误差。

相较于 DDQN 控制器, 传统 PI 控制方法的误差约为其 16、3、16 倍。综上所述, 当 CPL 功率发生改变时 RL 控制器相比 PI 控制器能更快达到稳定, 并且有着更小的超调量, 能够更好的实现母线电压快速跟踪。

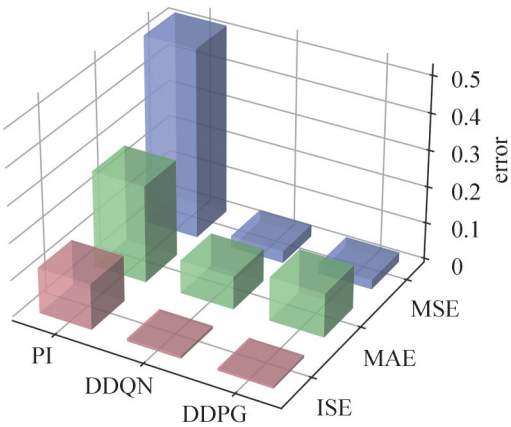


图 12 工况三下电压误差对比
Fig. 12 Comparison of voltage errors under working condition three

为了比较不同控制方法的瞬态性能, 将三种控制策略在面对三种工况改变时的最大输出电压上升/下降(MOVR/MOVD)在表 4 中给出。可以看出, 在面对三种不同扰动时, 强化学习控制器均能缩小瞬态压升/压降。其中 DDPG 控制器在三种工况中整体表现最好, 动态性能上, 拥有更小的电压波动与更快的稳定时间。静态性能上, 稳态误差小, 稳定后的电压抖动更小。

表 4 三种工况下最大输出电压升/降数值				
Table 4 Maximum output voltage rise/drop value under three working conditions				
工况	指标	PI	DDQN	DDPG
一	MOVR	0.119 8	0.034 3	0.013 7
	MOVD	0.296 1	0.016 7	0.014 1
二	MOVR	0.150 5	0.051 3	0.011 8
	MOVD	0.304 8	0.057 3	0.059 5
三	MOVR	1.368 0	0.886 5	0.246 8
	MOVD	4.731 0	1.396 0	1.001 5

5 结论

本文针对典型Boost变换器提出了一种基于深度强化学习的控制方法。首先基于两种强化学习算法设计了Boost变换器的控制器。基于Dymola平台搭建了仿真模型,并开发RLM工具包实现环境模型与强化学习智能体互连。最后进行了微源电压变化、阻性负载变化、恒功率负载变化三种扰动下的仿真比较验证,验证了RL控制器在面对各种扰动时的瞬态性能。仿真结果表明,与传统的双闭环PI控制方法相比,RL控制器均能以更小的超调在更短的时间内实现母线电压快速跟踪,多种衡量指标也印证了RL控制策略的优越性。

参考文献:

- [1] 李洪恩,胡健,唐志扬,等.孤岛直流微电网DC-DC变换器的非线性控制技术[J].电源学报,2020,18(2):4-14.
Li Hongen, Hu Jian, Tang Zhiyang, et al. Nonlinear Control Technology for DC-DC Converter in Islanded DC Microgrid[J]. Journal of Power Supply, 2020, 18(2): 4-14.
- [2] Ghosh S K, Roy T K, Pramanik A H, et al. Design of Nonlinear Backstepping Double-Integral Sliding Mode Controllers to Stabilize the DC-Bus Voltage for DC-DC Converters Feeding CPLs[J]. Energies (S1996-1073), 2021, 14(20): 6753.
- [3] Boutchich N, Moufid A, Bennis N, et al. A Constrained MPC Approach Applied to Buck DC-DC Converter for Greenhouse Powered by Photovoltaic Source[C]// 2020 International Conference on Electrical and Information Technologies (ICEIT). Piscataway, NJ, USA: IEEE, 2020: 1-6.
- [4] 陶彩霞,王伟斌,祝曦,等.带恒功率负载Buck变换器的模型预测控制[J].湖南大学学报(自然科学版),2021,48(8):96-102.
Tao Caixia, Wang Weibin, Zhu Xi, et al. A Model Predictive Control for Buck Converter with Constant Power Load[J]. Journal of Hunan University Natural Sciences, 2021, 48(8): 96-102.
- [5] 周雪松,李康,马幼捷.DC-DC变换器滑模变结构控制研究[J].电力系统及其自动化学报,2021,33(3):11-17.
Zhou Xuesong, Li Kang, Ma Youjie. Research on Sliding-Mode Variable-Structure Control of DC-DC Converter [J]. Proceedings of the CSU-EPSSA, 2021, 33(3): 11-17.
- [6] Fan Jingwen, Li Shihua, Wang Junxiao, et al. A GPI Based Sliding Mode Control Method for Boost DC-DC Converter[C]// 2016 IEEE International Conference on Industrial Technology (ICIT). Piscataway, NJ, USA: IEEE, 2016: 1826-1831.
- [7] Yfoulis C, Papadopoulou S, Voutetakis S. Robust Linear Control of Boost and Buck-Boost DC-DC Converters in Micro-grids with Constant Power Loads[J]. Energies (S1996-1073), 2020, 13(18): 4829.
- [8] 杨航,刘凌,阎治安,等.双闭环Buck变换器系统模糊PID控制[J].西安交通大学学报,2016,50(4):35-40,67.
Yang Hang, Liu Ling, Yan Zhian, et al. A Fuzzy PID Control Strategy for Buck Converter System of Double Closed Loop Circuits[J]. Journal of Xi'an Jiaotong University, 2016, 50(4): 35-40, 67.
- [9] Farsizadeh H, Gheisarnejad M, Mosayebi M, et al. An Intelligent and Fast Controller for DC/DC Converter Feeding CPL in a DC Microgrid[J]. IEEE Transactions on Circuits and Systems II: Express Briefs (S1549-7747), 2020, 67(6): 1104-1108.
- [10] Raj R N, Purushothaman K V, Singh N A. Adaptive TSK-Type Neural Fuzzy Controller for Boost DC-DC Converter[C]// 2017 IEEE International Conference on Circuits and Systems (ICCS). Piscataway, NJ, USA: IEEE, 2017: 441-446.
- [11] Khorashadizadeh S, Mahdian M. Voltage Tracking Control of DC-DC Boost Converter Using Brain Emotional Learning[C]// 2016 4th International Conference on Control, Instrumentation, and Automation (ICCIA). Piscataway, NJ, USA: IEEE, 2016: 268-272.
- [12] 贾振宇,刘子龙.一种通过强化学习的四旋翼姿态控制算法[J].小型微型计算机系统,2021,42(10):2074-2078.
Jia Zhenyu, Liu Zilong. Quadrotor Attitude Control Algorithm Based on Reinforcement Learning[J]. Journal of Chinese Computer Systems, 2021, 42(10): 2074-2078.
- [13] 董豪,杨静,李少波,等.基于深度强化学习的机器人运动控制研究进展[J].控制与决策,2022,37(2):278-292.
Dong Hao, Yang Jing, Li Shaobo, et al. Research Progress of Robot Motion Control Based on Deep Reinforcement Learning[J]. Control and Decision, 2022, 37(2): 278-292.
- [14] 张耀中,许佳林,姚康佳,等.基于DDPG算法的无人机集群追击任务[J].航空学报,2020,41(10):314-326.
Zhang Yaozhong, Xu Jialin, Yao Kangjia, et al. Pursuit Missions for UAV Swarms Based on DDPG Algorithm [J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(10): 314-326.
- [15] 熊珞琳,毛帅,唐漾,等.基于强化学习的综合能源系统管理综述[J].自动化学报,2021,47(10):2321-2340.

- Xiong Luolin, Mao Shuai, Tang Yang, et al. Reinforcement Learning Based Integrated Energy System Management: A Survey[J]. *Acta Automatica Sinica*, 2021, 47(10): 2321-2340.
- [16] 周毅, 周良才, 丁佳立, 等. 基于深度强化学习的电网拓扑优化及潮流控制[J]. *上海交通大学学报*, 2021, 55(增 2): 7-14.
- Zhou Yi, Zhou Liangcai, Ding Jiali, et al. Power Network Topology Optimization and Power Flow Control Based on Deep Reinforcement Learning[J]. *Journal of Shanghai Jiaotong University*, 2021, 55(S2): 7-14.
- [17] Hajihosseini M, Andalibi M, Gheisarnejad M, et al. DC/DC Power Converter Control-Based Deep Machine Learning Techniques: Real-Time Implementation[J]. *IEEE Transactions on Power Electronics* (S0885-8993), 2020, 35(10): 9971-9977.
- [18] Gheisarnejad M, Farsizadeh H, Tavana M R, et al. A Novel Deep Learning Controller for DC-DC Buck-Boost Converters in Wireless Power Transfer Feeding CPLs[J]. *IEEE Transactions on Industrial Electronics* (S0278-0046), 2021, 68(7): 6379-6384.
- [19] Lukianykhin O, Bogodorova T. ModelicaGym: Applying Reinforcement Learning to Modelica Models[C]// 9th International Workshop on Equation-Based Object-Oriented Modeling Languages and Tools. New York, NY, USA: Association for Computing Machinery, 2019: 27-36.