

基于深度强化学习的卫星电源冗余电池均衡控制策略

叶泽雨^{1,2}, 尹靖元^{1,2}, 贾海鹏³, 师长立^{1,2}, 韦统振^{1,2}, 罗彦¹

(1. 中国科学院电工研究所, 北京 100190; 2. 中国科学院大学电子电气与通信工程学院, 北京 100049;
3. 中国空间技术研究院钱学森空间技术实验室, 北京 100094)

摘要: 为提高卫星蓄电池组的智能化管理能力水平, 提出了一种基于深度强化学习的卫星冗余电池均衡控制策略. 训练智能体根据蓄电池组当前的运行状态给出动作, 改变单体电池的投入状态和数量, 实现单体之间容量均衡, 减小母线电压变化范围并减少开关调节次数. 在 MATLAB\Simulink 和 OpenAI 的 gym 环境中分别搭建了电池组仿真环境对智能体进行了训练, 通过算例检验了该策略的可行性并与基于阈值的控制方法进行了比较, 证明了方法可以有效的实现电池间均衡并减小母线电压的变化范围.

关键词: 电池管理; 冗余均衡; 可重构电池组; 深度强化学习; 电池容量均衡; 母线电压

基金项目: 国防基础科研计划资助(No.JCKY2019130C002)

中图分类号: TM912; TN710

文献标识码: A

文章编号: 0372-2112(2023)09-2419-09

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20211267

Balance Control Strategy of Redundant Battery in Satellite Power Supply Based on Deep Reinforcement Learning

YE Ze-yu^{1,2}, YIN Jing-yuan^{1,2}, JIA Hai-peng³, SHI Chang-li^{1,2}, WEI Tong-zhen^{1,2}, LUO Yan¹

(1. Institute of Electrical Engineering Chinese Academy of Sciences, Beijing 100190, China;

2. School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China;

3. China Academy of Space Technology Qian Xuesen Laboratory of Space Technology, Beijing 100094, China)

Abstract: In order to improve the intelligent management ability of satellite battery pack, this paper proposes a satellite redundant battery balancing control strategy based on deep reinforcement learning. The method can train the agent to extract the characteristics of the current operation state of the battery group and judge it, change the input state and quantity of the battery unit, realize the capacity balance between the battery units, reduce the range of bus voltage change and reduce the switching adjustment times. The battery pack simulation environment is built in MATLAB \ Simulink and OpenAI gym environment respectively to train the reinforcement learning agent. The feasibility of this method is verified by an example, and compared with the traditional sorting method, it is proved that this method can effectively achieve the capacity balance between batteries and reduce the variation range of bus voltage.

Key words: battery management; redundancy balance; reconfigurable battery pack; deep reinforcement learning; battery capacity balance; bus voltage

Foundation Item(s): National Defense Basic Scientific Research Program of China (No.JCKY2019130C002)

1 引言

卫星电源系统由太阳阵、电池组、电源控制器等部件组成^[1]. 在地影期由电池组独立提供能量, 因此确保电池组的可靠运行是卫星电源系统的关键问题之一^[2].

锂离子电池已经在卫星电源系统中得到广泛应

用. 锂电池单体电压较低且输出能力和容量有限, 通常采用大量单体串并联组成大容量电池组^[3]. 如果电池单体荷电状态(State Of Charge, SOC)不一致, 容量较小的单体可能过充/放, 将加速电池老化^[4,5], 严重影响其使用寿命和安全性. 与地面电池组相比, 卫星电源电池

组还存在如下问题:(1)锂电池端电压随容量变化较大,为了提高供电效率和响应时间,电源系统往往需要将电池组直挂母线上^[6],这难以对母线电压进行控制;(2)锂电池单体故障率较高,相较于一般电池组,卫星电源系统无法更换故障单体;(3)卫星电源系统对体积和重量有较高的要求,均衡电路设计需要简单可靠。

为了解决上述问题,需要在串联电池组中增加冗余电池,并对电池组输出电压进行控制.在单体电池损坏后,可以切除故障单体,提高系统容错能力.电池组还可以利用冗余电池实现单体间的容量均衡,不需要额外的均衡电路.在系统工作过程中,还可以改变投入的串联电池数量,改变电池组端口电压。

针对含有冗余电池的串联电池组的容量管理方法,文献[7]将冗余电池作为附加电源,实现非冗余电池容量均衡.但是该方法不能控制冗余电池的容量,在冗余电池充满或耗尽时将失去均衡能力.文献[8]提出了一种以电池单体电压和SOC偏差为依据的均衡方法.文献[9]提出了一种等能量变化/时间切换的电池均衡策略,并讨论了冗余电池数量、切换频率和效率等因素的影响.文献[10]提出了一种基于冒泡排序的均衡方式,并与升压变换器配合使用,提升了电池组的容错能力.这些方法通过确定的阈值根据电池的剩余容量进行控制,存在着均衡效果和切换次数的矛盾.文献[11]提出了一种基于二次规划求解单体电池最优分配时间比例的均衡方法.该方法需要已知充/放电周期使用电量和单体的效率。

由于锂电池端电压随容量变化较大,母线电压会有较大波动,减少母线电压的变化范围有利于降低二次电源的设计要求.改变串联电池数量可以减少串联电池组端电压的变化范围,电池单体两端的电压受充放电电流影响,剩余容量相同的单体在静置和工作状态下的端电压有较大差别,不易判断电池节数改变的时间点.随着串联电池节数与冗余电池节数的增加,均衡算法的实现也会更加困难。

近年来,人工智能技术迅速发展,已经广泛应用到多个行业.与经典的优化方法、基于规划的方法和启发式算法相比,强化学习方法具有不需要依赖明确的目标函数、容易收敛、能够进行实时决策等优点^[12].本文提出一种基于强化学习的串联电池控制策略.利用冗余电池均衡电池容量并动态调节串联电池数量,在尽可能保证直流电压稳定的同时减少开关动作次数.在MTALB\Simulink和OpenAI的gym环境中完成模型搭建和强化学习智能体训练,最后进行了算例检验并在Jetson Xavier NX处理器部署了算法,验证了所提出的强化学习方法对电池均衡控制的有效性和实时性。

2 系统建模

2.1 卫星电源系统及均衡拓扑

卫星电源系统主要由太阳阵、太阳阵控制器和电池组构成.为了简化结构,提高系统效率,一些卫星电源系统将电池组直接挂在母线上^[13],如图1所示.在地影期,母线电压完全取决于电池组电压,电池电压变化会直接影响供电电压。

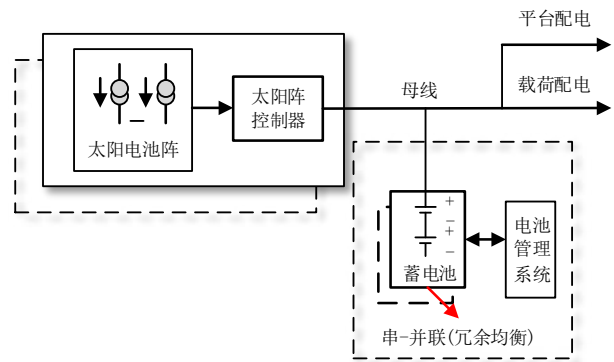


图1 卫星电源系统结构图

针对图1中蓄电池组,本文采用的串联电池电路的拓扑结构如图2所示^[8].其中 $B_i(i=1,2,\dots,N)$ 为电池单体,每节电池配备一个串联开关 K_i 和一个并联开关 K'_i .当 K_i 闭合, K'_i 断开时,单体电池投入; K_i 断开, K'_i 闭合时,单体电池切除.为了避免两个开关同时闭合引起短路,应选用常开式开关,并配备熔断器。

蓄电池组的串联冗余均衡方式下,如果电池串联数量为 N 节,备用电池数量为1节.在额定工作情况下,只有 $N-1$ 节电池串联工作.在光照期对电池充电时,如果检测到 B_i 的容量较大而其他电池的容量较小时,将 B_i 旁路并将冗余电池投入,保证串联电池总数仍然为 $N-1$ 节;同样,在地影期,检测到 B_i 的容量较小,而其它电池的容量较大时,将 B_i 旁路,由其他电池为负载提供能量。

2.2 电池组建模

电池的SOC定义为:

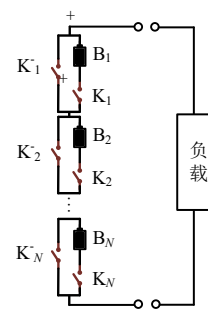


图2 冗余均衡拓扑

$$SOC = \frac{Q_t}{Q_0} \quad (1)$$

其中, Q_t 为电池在 t 时刻的剩余容量, Q_0 为电池额定容量, 即电池充满后可放出的最大容量。

电池模型可以等效为受控电压源 E 和电阻 R 串联,

$$V_{BAT} = E - R \cdot i \quad (2)$$

其中, V_{BAT} 为电池两端电压, i 为流经电池单体的电流, $i > 0$ 为单体放电, $i < 0$ 为单体充电。

电源系统需要满足电流平衡条件。在光照期太阳阵为负载供电, 多余的能量为蓄电池充电, 如果光照不足或负载需求过大, 则由太阳阵和蓄电池同时为负载供电。在地影期, 负载由蓄电池供电。

$$I_{PV} + I_{BAT} = I_{LOAD} \quad (3)$$

其中, $I_{BAT} > 0$ 为电池放电, $I_{BAT} < 0$ 为电池充电。

由于电池的容量和充放电能力有限, 需要满足电流约束, 容量约束和工作电压约束。

$$I_{BAT_MIN} \leq I_{BAT} \leq I_{BAT_MAX} \quad (4)$$

$$SOC_{MIN} \leq SOC \leq SOC_{MAX} \quad (5)$$

$$V_{BAT_MIN} \leq V_{BAT} \leq V_{BAT_MAX} \quad (6)$$

3 深度强化学习算法

3.1 基本架构

传统的控制方法通过控制器对被控对象发出控制信号, 再由反馈对控制信号进行修正, 控制策略通常由数值或逻辑运算组成, 需要设计者具备足够的专业知识, 如图 3 所示。深度学习的特点是不需要先验知识, 可以通过海量数据提取特征进行自主学习, 但是缺乏决策能力, 而强化学习具有决策能力, 却需要人工提取特征。将两者进行结合为解决复杂系统的决策问题提供了新的思路^[14-17]。

强化学习的问题模型通常被表示为一个元组 (S, A, P, R) , 其中, S 表示状态的集合, A 表示动作的集合,

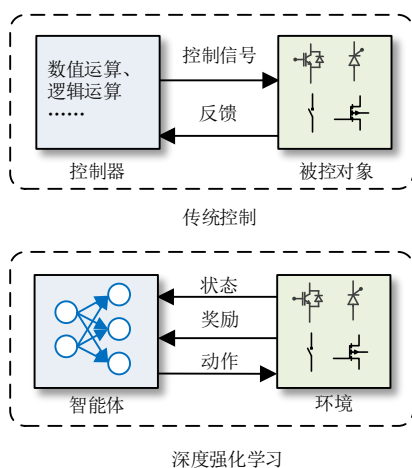


图3 深度强化学习和传统控制方法的区别

P 表示状态转移概率, R 表示奖励函数。

强化学习通过代理观察环境的当前状态 $s \in S$, 并根据状态决定要采取的动作 $a \in A$ 。一种状态可能对应多种动作, 每种动作有不同的选择概率 p , p 由策略 π 决定。

$$\pi(a|s) = p[A_t = a | S_t = s] \quad (7)$$

动作对状态产生改变后, 会生成奖励。强化学习的价值由当前奖励和折扣的未来奖励组成。如果该奖励是正面的, 则随后执行该奖励的概率会增加。在状态 s 中, 动作 a 对应的价值如式(8)所示:

$$Q_\pi(s, a) \equiv E[R_1 + \gamma R_2 + \dots | S_0 = s, A_0 = a, \pi] \quad (8)$$

式中, $\gamma \in [0, 1]$ 为折扣因子。使用贪婪策略可以选择当前价值最大的动作, 最优价值为:

$$Q_*(s, a) = \max_\pi Q_\pi(s, a) \quad (9)$$

实际中许多问题难以用模型描述, 状态转移概率 P , 以及每一个动作或状态对应的即时奖励可能是未知的。 Q 学习 (Q-learning) 是一种不基于环境模型、基于价值的强化学习算法。其主要思路是定义状态动作价值函数, 即 Q 函数。在随机初始化参数后, 将观测数据代入到式(10)中对 Q 函数进行迭代学习。

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left(r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right) \quad (10)$$

其中, s_t 和 s_{t+1} 分别为 t 时刻的状态和 $t+1$ 时刻的状态, a_t 为 t 时刻的动作, α 为学习率。

如果状态和动作有限, Q 学习方法可以生成一个表格。但是如果状态或动作连续, 使用表格表示策略就不可行了。使用参数化的动作价值函数 $Q(s, a | \theta)$ 作为函数逼近器, 可以处理连续的状态。更新方法是梯度下降法, 使当前价值 Q 接近目标值 Y_t^Q 。

$$\theta_{t+1} = \theta_t + \alpha (Y_t^Q - Q(S_t, A_t; \theta_t)) \nabla_{\theta_t} Q(S_t, A_t; \theta_t) \quad (11)$$

其中目标值 Y_t^Q 如式(12)所示:

$$Y_t^Q \equiv R_{t+1} + \gamma Q(S_{t+1}, \max_a Q(S_{t+1}, a; \theta_t); \theta_t) \quad (12)$$

在深度 Q 网络 (Deep-Q-Network, DQN) 方法中, 使用神经网络代替价值函数, 并增加了目标网络, 参数为 θ^- 。其中 θ^- 每隔几步由在线网络的参数 θ 复制过来。同时 DQN 使用了经验回放的训练方法, 可以使训练容易收敛且更稳定。DQN 的目标值如式(13)所示:

$$Y_t^{DQN} \equiv R_{t+1} + \gamma Q(S_{t+1}, \max_a Q(S_{t+1}, a; \theta_t^-); \theta_t^-) \quad (13)$$

与式(13)对比可知, DQN 中的贪婪策略与 Q 值更新都是在目标网络中执行的。

由于在值函数逼近中, 动作选择和价值计算使用的参数是一样的, 计算最大值的步骤会使估值偏大。双深度 Q 网络 (Double Deep-Q-Network, DDQN) 算法^[18] 在 DQN 的基础上进行了改进, 其更新方法与 DQN 一样, 但

是目标值替换为

$$Y_t^{\text{DoubleDQN}} \equiv R_{t+1} + \gamma Q\left(S_{t+1}, \arg \max_a Q(S_{t+1}, a; \theta_i); \theta_i^-\right) \quad (14)$$

与 DQN 的目标值进行对比可以看出, DDQN 的贪婪策略在在线网络中执行, 而 Q 值更新在目标网络中执行. 由于两个网络之间存在一定的更新间隔, 动作选择和价值计算被分开, 可以使价值计算更加准确、加速训练过程并更容易收敛.

3.2 控制目标

本文需要解决的问题包括串联电池组的均衡以及电池组端口电压调节.

首先是串联电池组的均衡, 电池组的均衡程度可由式(15)计算.

$$\Delta \text{SOC} = \frac{\text{SOC}_{\text{BAT}_{\max}} - \text{SOC}_{\text{BAT}_{\min}}}{|\text{SOC}_{\text{END}} - \text{SOC}'|} \quad (15)$$

其中, $\text{SOC}_{\text{BAT}_{\max}}$ 为当前所有电池中 SOC 最大的电池单体的 SOC, $\text{SOC}_{\text{BAT}_{\min}}$ 为当前所有电池中 SOC 最小的电池单体的 SOC, SOC' 为所有电池单体的平均 SOC, SOC_{END} 为电池充电或放电的终点. 在充电阶段, $\text{SOC}_{\text{END}} = \text{SOC}_{\text{MAX}}$; 在放电阶段, $\text{SOC}_{\text{END}} = \text{SOC}_{\text{MIN}}$. 这种设置对电池 SOC 偏离程度要求在充/放电初始阶段较小, 在充/放电过程中不断增加.

其次, 需要尽可能减少母线电压变化范围. 母线电压的波动范围可由式(16)计算.

$$\Delta V = \frac{V_B - V_R}{V_R} \quad (16)$$

其中, V_B 为母线电压的实际值, V_R 为母线电压的额定值.

投入和切除电池单体需要对开关进行切换, 频繁的通断开关, 会减少开关的使用寿命, 在电池的离散程度可以接受的情况下, 应尽可能的减少电池单体的切换次数. 开关的动作次数 F 可由式(17)计算.

$$F = \sum_i^N |a_i^{t+1} - a_i^t| \quad (17)$$

式中 N 为串联电池总节数, a_i^t 为 t 时刻开关 K_i 的动作, 其中 1 为接入, 0 为切除, a_i^{t+1} 为 $t+1$ 时刻开关 K_i 的动作, 当改变开关动作时两者差值的绝对值为 1, 不改变时为 0.

在实现串联电池均衡与母线电压调节时, 需要改变电池单体的工作状态, 控制变量为开关的通断情况. 动作为开关的组合状态.

$$a = (a_1, a_2, \dots, a_N) \quad (18)$$

式中 a_i 为 1 或 0. 其中, 1 表示电池投入, 0 表示电池切除.

3.3 训练过程

本文所使用的强化学习算法的整体框架图如图 4

所示, 包括环境、状态、奖励、动作等要素. 在训练过程中, 强化学习智能体通过模拟环境进行自主学习, 将每一个控制周期获得的状态、奖励和动作等信息存入记忆库中, 并从记忆库中抽样对神经网络更新.

当前环境的状态包括系统的母线电压 V_B , N 节电池的 SOC, N 节电池的电压 V_{BAT_j} , 控制 N 节电池的 N 个开关的上一个控制周期的动作信号 a_i . 动作空间如式(18)所示, 为所有开关状态的组合. 奖励包括开关状态变化的惩罚, 母线电压偏离期望电压的惩罚和电池 SOC 偏差惩罚, 如式(19)所示.

$$R = \alpha_1(\Delta V - T_1) + \alpha_2 \sum_i^N |a_i^{t+1} - a_i^t| + \alpha_3(\Delta \text{SOC} - T_2) \quad (19)$$

这里 T_1 和 T_2 分别为母线电压和电池 SOC 偏离程度的阈值.

强化学习的训练流程如图 5 所示. 其中, t 为采样周期, T 为仿真时间, j 为当前训练次数, J 为训练总次数. 当 $\text{SOC}_{\text{BAT}_{\min}}$ 小于 SOC_{MIN} 时提前结束本次训练.

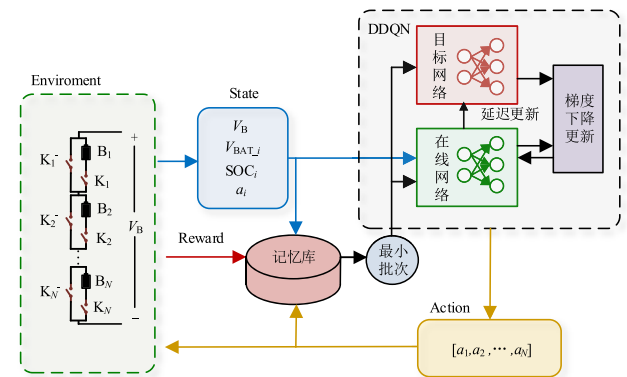


图4 算法整体框架图

4 算例分析

4.1 仿真设置

电池电压的工作范围为 2.6~4.2 V, 标称容量为 3 000 mAh. 母线电压的额定值设置为 28 V. 本算例中智能体需要获得的信息包括系统的母线电压 V_B , 9 节电池的 SOC: $[\text{SOC}_1, \text{SOC}_2, \dots, \text{SOC}_9]$, 9 节电池的电压: $[V_{\text{BAT}_1}, V_{\text{BAT}_2}, \dots, V_{\text{BAT}_9}]$, 控制 9 节电池的 9 个开关的上一个控制周期的动作信号: $[a_1, a_2, \dots, a_9]$.

锂离子电池电压变化范围较大, 为了减少母线电压的变化范围, 在工作过程中灵活选择同时串联的电池数量. 设置 46 个动作组合, 包括 9 节电池全部投入 ($C_9^9=1$), 投入 8 节电池 ($C_9^8=9$) 和 7 节电池 ($C_9^7=36$) 三种情况.

考虑到冗余电池的存在, 需要控制电池进行轮流使用, 所以在充放电过程的前中期并不需要电池完全均衡, 而是在整个放电周期结束之前逐渐均衡即可, 电

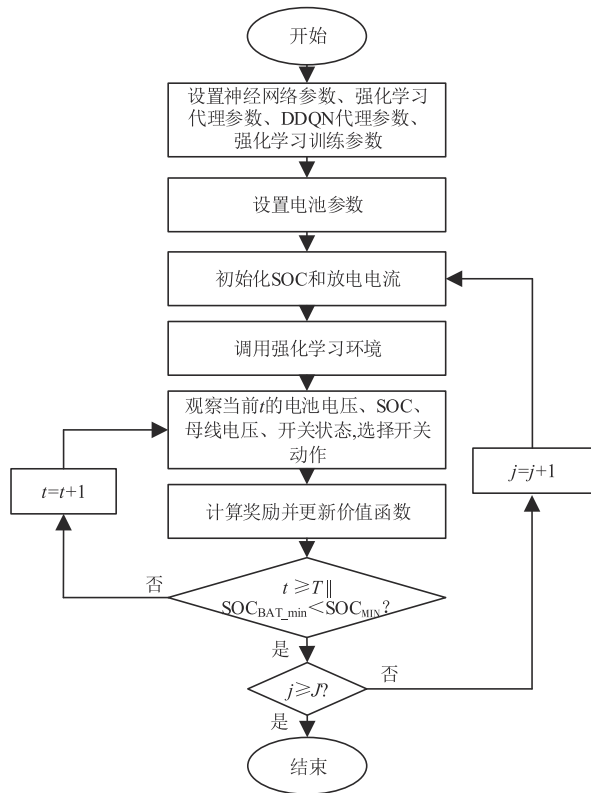


图5 深度强化学习训练流程

池偏离程度的计算式(15)很好的体现了这一点. 对于式(19), 设定当母线电压的偏离程度超过5%, SOC 偏离程度超过10%时增加惩罚, 同时对每一步的开关动作次数增加惩罚.

DDQN算法中的神经网络设置为四层, 其中输入层共有28个节点, 对应28个观察信号, 隐藏层1共有112个节点, 隐藏层2共有184个节点, 输出层共有46个节点, 对应46个动作.

由于电池充电阶段母线电压会受到太阳阵控制器工作模式的影响. 这里只对电池放电阶段进行仿真. 仿真时间设置为30 min, 强化学习的采样设置为60 s/次. 设置电池初始SOC为80~100的随机数. 电池放电电流为5.5~7 A之间的一个随机数.

4.2 结果分析

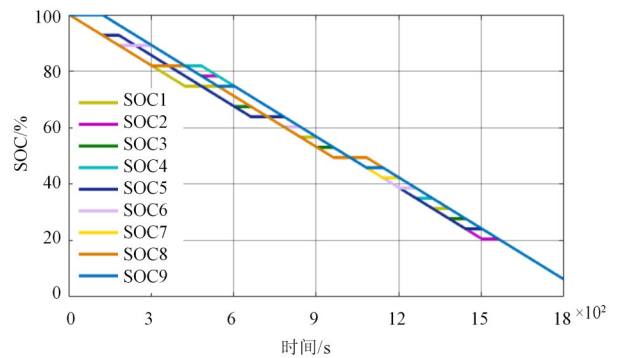
为了检验电池均衡和电压控制效果, 在电池初始条件均衡和不均衡两种条件下进行了仿真实验.

初始条件均衡时, 初始SOC设置为100. 放电电流设置为6.5 A. 图6(a)为各节电池放电过程中所有电池单体SOC变化曲线. 图6(b)为各电池单体SOC的极差值, 从图6(b)中可以看出, 在电池组放电过程中, 单体电池间SOC的最大差异在7左右. 在放电结束时, SOC的差异减小到0. 图6(c)为母线电压变化情况, 图6(d)为投入电池节数. 在起始阶段, 单体电池电压较高, 投入电池节数为7节, 随着单体电池电压的升高, 为了维

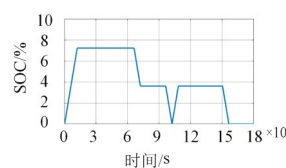
持母线电压接近28 V, 投入电池节数依次增加为8节、9节, 电压变化范围在23.9~30.3 V左右.

图6(e)为每个控制周期开关的动作次数, 表示在当前控制周期共有多少节电池的工作状态发生了变化. 从图中可以看出, 电池的投入和切除并不是时刻进行的, 通过对开关动作设置惩罚, 可以在满足控制目标的同时, 减少开关动作次数.

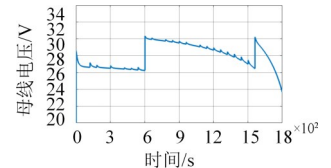
初始条件不均衡的仿真中, 初始SOC分别为100, 99, 95, 91, 90, 89, 85, 81和80, 放电电流为5.8 A.



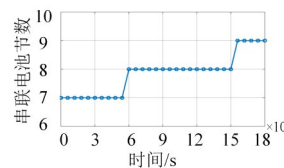
(a) 电池SOC变化曲线



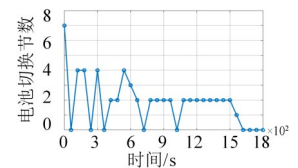
(b) 电池SOC极差



(c) 母线电压变化曲线



(d) 串联电池数量



(e) 电池切换节数

图6 初始条件均衡控制效果

放电过程中电池SOC的变化曲线如图7(a)所示, 图7(b)为放电过程中电池组SOC的极差值, 各节电池SOC的最大值和最小值的差值为20, 在800 s时, 电池的SOC差值减少到5以下, 结束时差值在3以下. 在放电过程中实现了SOC均衡. 母线电压的变化范围如图7(c)所示, 在500 s之前, 接入系统的电池节数为7节. 随着单节电池电压降低, 接入的电池数增加为8节和9节. 系统电压也维持在接近28 V. 可以看出该控制策略在实现电池均衡的同时并没有增大系统母线电压的变化范围.

在现有的针对可重构电池组的研究中, 文献[7, 8,

10]均是基于设置电池单体间不均衡程度的阈值来实现均衡控制的,基本对结构如图8所示.

该类方法利用排序与阈值判断,从 N 节电池中挑选 $N-1$ 节同时工作.首先对所有电池单体的容量进行排序确定 SOC_{BAT_max} 和 SOC_{BAT_min} ,然后使用式(16)判断是否达到均衡条件,如果未达到,则保持上一个控制周期的开关状态;如果达到,则将SOC最大的电池单体的控制开关置0,并将冗余电池单体的开关置为1.对比结果中的控制周期、均衡阈值设置以及仿真设置与上一小节中初始条件不均衡的设置一致.

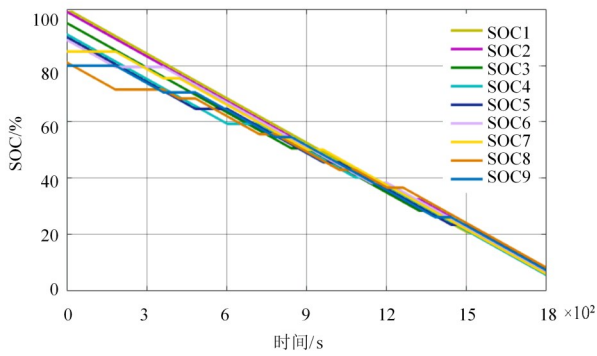


图8 基于阈值的均衡控制方法

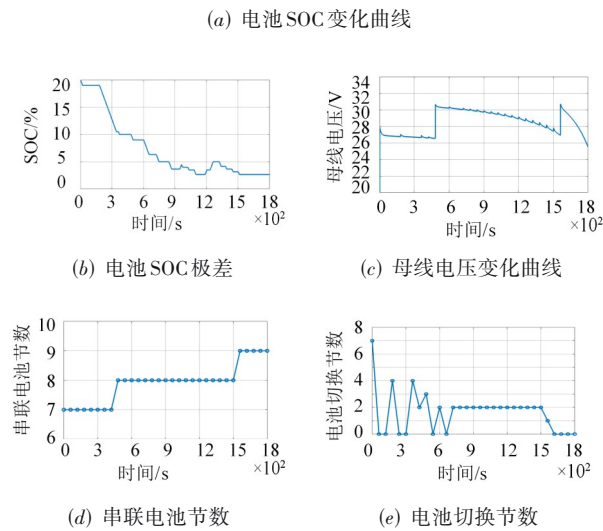


图7 初始条件不均衡控制效果

对比方法的控制效果如图9所示,其中电池单体SOC的变化趋势如图9(a)所示,从图9(b)可以看出,该方法的最终均衡目标与图7(b)相同.从图9(b)中可以看出,由于没有调整工作电池的数目,母线电压的变化范围为31~22 V,在图7(b)中母线电压的变化范围为30.75~25.5 V,通过改变电池节数,母线电压的变化减少了56%.从图9(d)中可以看出,如果不考虑初始变化,开关动作次数为56次,在图7(e)中,使用强化学习方法的开关动作次数为50次,开关次数也有所减少.

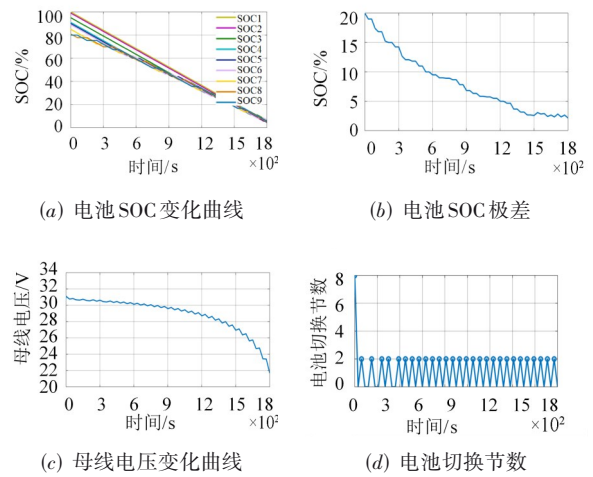


图9 常规方法的控制效果

4.3 参数对比

为了研究不同参数对实验结果的影响,本文对DDQN算法框架的不同优化器,包括rmsprop、adam和sgdm,以及关键的超参数学习率(Learning Rate, lr)和批大小(Batchsize, bs)进行了比较.设置了5个参数组合,不同参数组合下lr和bs的设置如表1所示,其他参数的设置如表2所示.图10(a)~(c)分别是使用rmsprop、adam和sgdm优化器在不同参数组合下训练过程的平均奖励曲线,奖励越高,说明当前控制策略的效果越接近预期.

对比三个优化器的训练过程可以看出,使用adam

表1 对比参数

参数	数值
折扣率	0.9
记忆库大小	100 000
梯度阈值	2
目标平滑系数	1
目标更新频率	4

表2 其他参数

组合	学习率	批大小
参数组合1	0.005	128
参数组合2	0.001	64
参数组合3	0.001	128
参数组合4	0.001	256
参数组合5	0.000 5	128

和 rmsprop 优化器训练 5 000 次后在 lr 为 0.000 5 和 0.001 时都可以收敛到较好的结果,而采用 sgdm 优化器的训练结果几乎都没有收敛到另外两个优化器的

最佳结果. 在已有的参数对比中,采用 adam 优化器的收敛速度要略快于 rmsprop. 在 adam 和 rmsprop 优化器的训练结果中,不同参数下的训练过程平均奖励的变化趋势相同,当 lr 为 0.005 时可以更快收敛,但是训练效果较差. 而 lr 过小(0.000 5)会影响收敛速度. 通过比较参数组合 2、3 和 4 可知,增大 bs 可以适当提升收敛速度,使用 adam 和 rmsprop 优化器的训练结果中,参数组合 4 都是最先收敛到最优值的,但是考虑到增大 bs 也会增加计算负担,在收敛速度差距不大的情况下 bs 也不宜过大.

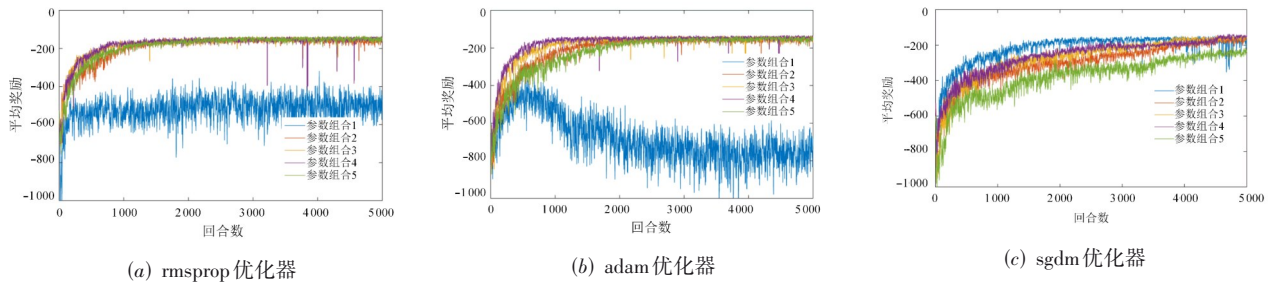


图 10 不同优化器和参数的奖励曲线

4.4 实验验证

为了进一步验证方法的时效性和可行性,制作了英伟达 Jetson Xavier NX 处理器载板,对算法进行了验证,Jetson Xavier NX 是一款基于 arm 的嵌入式 GPU 硬件平台,可以部署人工智能算法.

首先,使用电池测试仪对松下 18650 型电池进行放电测试,如图 11 所示. 不同电流大小下的放电曲线如图 12 所示. 根据电池放电曲线,建立了电压源和电阻串联的电池模型,电池内阻根据数据手册设定为 0.04 Ω,开路电压 E 与 SOC 的关系根据电池放电曲线拟合,如式 (20) 所示.

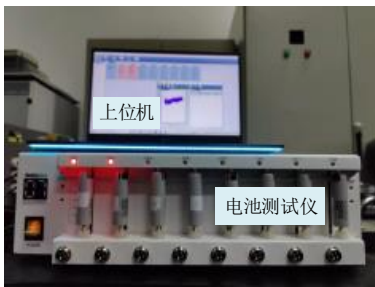


图 11 电池测试平台

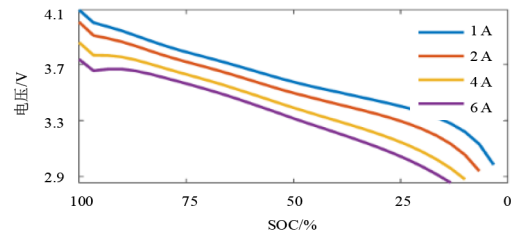


图 12 电池放电曲线

然后,使用 python 在 OpenAI 的强化学习环境库 gym 中搭建了本方法需要的强化学习环境,神经网络框架采用 pytorch,强化学习算法的训练参数与神经网络参数与 matlab 仿真一致. 最后,将训练好的模型和搭建的 gym 电池环境分别部署到两个 Jetson Xavier NX 上,信号通过搭配的 FPGA 使用 PCIE (Peripheral Component Interconnect Express) 传递,两个载板的 FPGA 通过 LVDS (Low-Voltage Differential Signaling) 通信,如图 13 所示.

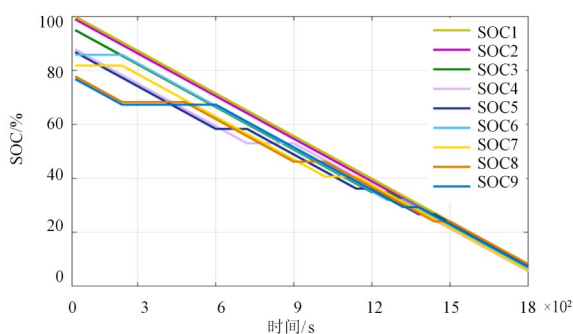
在程序运行的同时记录观察状态和执行动作,本算例的运行结果为如图 14 所示. 从结果可以看出,电池单体的 SOC 变化和电池组电压变化曲线和 Matlab 仿真中的结果基本一致,同时完成了电池单体均衡和电池组电压调节两个目标. 进行本算例 30 次控制的总时间为 0.84 s,完全可以满足 60 s/次的实时控制要求.

$$E = p_1 \times SOC^5 + p_2 \times SOC^4 + p_3 \times SOC^3 + p_4 \times SOC^2 + p_5 \times SOC + p_6 \tag{20}$$

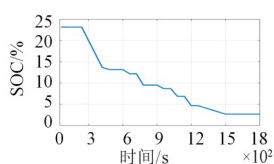
其中, p_1 至 p_6 为系数, $p_1=1.445 \times 10^{-9}$, $p_2=-4.06 \times 10^{-7}$, $p_3=4.3 \times 10^{-5}$, $p_4=-0.002 1$, $p_5=0.054$, $p_6=2.8$.



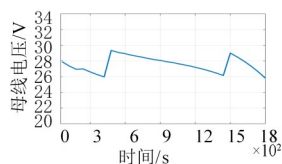
图 13 实验平台



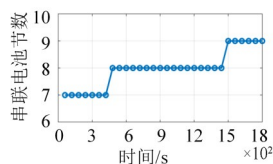
(a) 电池 SOC 变化曲线



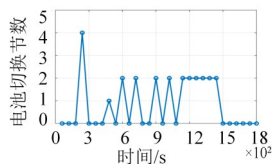
(b) 电池 SOC 极差



(c) 母线电压变化曲线



(d) 母线电压变化曲线



(e) 电池切换节数

图 14 实验结果

5 结论

针对卫星电源系统中串联电池组均衡和母线电压变化范围较大的问题,本文提出了一种基于深度强化学习的串联电池冗余均衡控制策略.主要创新点为:

(1)在串联电池组冗余均衡拓扑中,通过控制电池的投入和切除在实现电池容量均衡的同时改变串联电池节数调节母线电压.

(2)提出了基于 DDQN 强化学习的电池组控制算

法,设计强化学习算法的动作空间和奖励函数.

(3)使用 simulink 和 gym 分别搭建了电池组的强化学习环境,并在嵌入式硬件平台上进行了算例验证.

在算例分析中,本方法在满足电池均衡目标要求的同时减少了 56% 的母线电压变化范围和 11% 开关次数,并达到同样的均衡效果.该方法有助于实现高可靠电池管理,提升卫星电源系统的性能和安全性.

参考文献

- [1] 李朋,周军,于晓洲,等.翱翔之星电源系统的设计与在轨验证[J].太阳能学报,2018,39(4):1002-1007.
LI P, ZHOU J, YU X Z, et al. Design and on-orbit validation of electric power system for aoxiang-sat[J]. Acta Energetica Solaris Sinica, 2018, 39(4): 1002-1007. (in Chinese)
- [2] 张怀峰,江婧,张香燕,等.面向卫星电源系统的一种新颖异常检测方法[J].宇航学报,2019,40(12):1468-1477.
ZHANG H F, JIANG J, ZHANG X Y, et al. Novel anomaly detection method for satellite power system[J]. Journal of Astronautics, 2019, 40(12): 1468-1477. (in Chinese)
- [3] 赵汉宾,马锋,叶晓东,等.基于改进型飞渡电容的电池组动态均衡控制方法研究[J].电子学报,2018,46(4):819-826.
ZHAO H B, MA F, YE X D, et al. Battery dynamic balance control algorithm based on modified flying-capacitor[J]. Acta Electronica Sinica, 2018, 46(4): 819-826. (in Chinese)
- [4] 刘月峰,赵光权,彭喜元.多核相关向量机优化模型的锂电池剩余寿命预测方法[J].电子学报,2019,47(6):1285-1292.
LIU Y F, ZHAO G Q, PENG X Y. A lithium-ion battery remaining using life prediction method based on multi-kernel relevance vector machine optimized model[J]. Acta Electronica Sinica, 2019, 47(6): 1285-1292. (in Chinese)
- [5] 刘红锐,李园专,陈仕龙,等.锂离子蓄电池组充放电均衡器及均衡策略[J].电子学报,2016,44(8):1981-1987.
LIU H R, LI Y Z, CHEN S L, et al. Lithium-ion battery charging and discharging equalizer and balancing strategies [J]. Acta Electronica Sinica, 2016, 44(8): 1981-1987. (in Chinese)
- [6] 朱立颖,刘治钢,张晓峰.国外小天体探测器电源系统设计分析[J].航天器工程,2018,27(1):130-136.
ZHU L Y, LIU Z G, ZHANG X F. Analysis on power system design of probe for small celestial body[J]. Spacecraft Engineering, 2018, 27(1): 130-136. (in Chinese)
- [7] 廖力,纪锋,吴铁洲,等.采用附加电源的均衡电路与容量自均衡方法[J].电测与仪表,2020,57(7):48-53.
LIAO L, JI F, WU T Z, et al. Equalization circuit with ad-

- ditional power supply and capacity self-balancing method [J]. *Electrical Measurement & Instrumentation*, 2020, 57(7): 48-53. (in Chinese)
- [8] MANENTI A, ABBA A, MERATI A, et al. A new BMS architecture based on cell redundancy[J]. *IEEE Transactions on Industrial Electronics*, 2011, 58(9): 4314-4322.
- [9] 王莹, 严晓, 林文魁, 等. 具备冗余单元的电池组放电效率研究[J]. *电源技术*, 2018, 42(12): 1832-1834, 1926. WANG Y, YAN X, LIN W K, et al. Research on discharge efficiency of battery pack with redundant cells[J]. *Chinese Journal of Power Sources*, 2018, 42(12): 1832-1834, 1926. (in Chinese)
- [10] 王友仁, 黄薛, 耿星, 等. 容错航空蓄电池电源及其均衡管理[J]. *航空学报*, 2018, 39(5): 147-157. WANG Y R, HUANG X, GENG X, et al. Fault-tolerant battery power supply for aircraft and its active equalization management[J]. *Acta Aeronautica et Astronautica Sinica*, 2018, 39(5): 147-157. (in Chinese)
- [11] 李结胜. 高效高精度电池管理系统的设计和开发[D]. 武汉: 华中科技大学, 2019. LI J S. Design and Development of High-Efficiency and High-Precision Battery Management System[D]. Wuhan: Huazhong University of Science and Technology, 2019. (in Chinese)
- [12] 张自东, 邱才明, 张东霞, 等. 基于深度强化学习的微电网复合储能协调控制方法[J]. *电网技术*, 2019, 43(6): 1914-1921. ZHANG Z D, QIU C M, ZHANG D X, et al. A coordinated control method for hybrid energy storage system in microgrid based on deep reinforcement learning[J]. *Power System Technology*, 2019, 43(6): 1914-1921. (in Chinese)
- [13] 雷英俊, 马亮, 李海津. 一种小型 SAR 卫星电源系统设计[J]. *航天器工程*, 2019, 28(5): 101-104. LEI Y J, MA L, LI H J. Design of power supply system for small SAR satellite[J]. *Spacecraft Engineering*, 2019, 28(5): 101-104. (in Chinese)
- [14] TREMBLAY O, DESSAINT L A. Experimental validation of a battery dynamic model for EV applications[J]. *World Electric Vehicle Journal*, 2009, 3(2): 289-298.
- [15] 赵冬斌, 邵坤, 朱圆恒, 等. 深度强化学习综述: 兼论计算机围棋的发展[J]. *控制理论与应用*, 2016, 33(6): 701-717. ZHAO D B, SHAO K, ZHU Y H, et al. Review of deep reinforcement learning and discussions on the development of computer Go[J]. *Control Theory & Applications*, 2016, 33(6): 701-717. (in Chinese)
- [16] 杨启萌, 禹龙, 田生伟, 等. 基于深度强化学习的维吾尔语人称代词指代消解[J]. *电子学报*, 2020, 48(6): 1077-1083. YANG Q M, YU L, TIAN S W, et al. Anaphora resolution of uyghur personal pronouns based on deep reinforcement learning[J]. *Acta Electronica Sinica*, 2020, 48(6): 1077-1083. (in Chinese)
- [17] 冀大雄, 方文巍, 朱华, 等. 基于相对测量的水下机器人主动定位方法研究[J]. *电子学报*, 2021, 49(7): 1249-1256. JI D X, FANG W W, ZHU H, et al. Active localization of autonomous underwater vehicle using noisy relative measurement[J]. *Acta Electronica Sinica*, 2021, 49(7): 1249-1256. (in Chinese)
- [18] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q -learning[C]//Proceedings of the the Thirtieth AAAI Conference on Artificial Intelligence. Phoenix: AAAI Press, 2016: 2094-2

作者简介



叶泽雨 男, 1995年6月出生于安徽省淮南市. 现为中国科学院大学/中国科学院电工研究所博士研究生. 主要研究方向为深度学习、电力电子与电力传动.
E-mail: 451919724@qq.com



尹靖元(通讯作者) 男, 1987年10月出生, 于吉林省公主岭市. 中国科学院电工研究所副研究员, 硕士生导师. 主要从事源网荷储智能算法研究.
E-mail: yinjingyuan@mail.iee.ac.cn

贾海鹏 男, 1987年2月生于山西省太原市. 现为中国空间技术研究院钱学森空间技术实验室高级工程师. 从事柔性直流输电技术、人工智能航天应用等方面工作.
E-mail: jiahaipeng@qxslab.cn

师长立 男, 1984年3月出生于河北省衡水市, 现为储能系统工程师, 博士研究生. 主要研究方向为储能关键技术.
E-mail: shichangli@mail.iee.ac.cn

韦统振 男, 1976年8月出生于山东省单县, 博士. 现为中国科学院电工研究所研究员, 博士生导师. 主要从事直流电力设备、电力电子在电力系统中的应用研究.
E-mail: tzwei@mail.iee.ac.cn

罗彦 女, 1977年11月出生于山西省太原市, 硕士. 主要研究方向为人工智能及其应用.
E-mail: luoyan@mail.iee.ac.cn