

# 控制与决策

Control and Decision



## 阴影条件下基于迁移强化学习的光伏系统最大功率跟踪

杨博, THIDAR Swe, 钟林恩, 束洪春, 张孝顺, 余涛

引用本文:

杨博, THIDAR Swe, 钟林恩, 等. 阴影条件下基于迁移强化学习的光伏系统最大功率跟踪[J]. 控制与决策, 2020, 35(12): 2939–2949.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.0412>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

### 参数未知的离散系统Q-学习优化状态估计与控制

Q-learning optimal state estimation and control for discrete systems with unknown parameters

控制与决策. 2020, 35(12): 2889–2897 <https://doi.org/10.13195/j.kzyjc.2019.0180>

### 基于操作风险的双模式传感器管理方法

A dual-mode sensor management method based on operational risk

控制与决策. 2020, 35(12): 2993–2998 <https://doi.org/10.13195/j.kzyjc.2018.1541>

### 基于搜索空间划分与Canopy K-means聚类的种群初始化方法

Population initialization based on search space partition and Canopy K-means clustering

控制与决策. 2020, 35(11): 2767–2772 <https://doi.org/10.13195/j.kzyjc.2019.0358>

### 复合类别航站楼分配问题的改进和声搜索算法

Solving composite airport gate allocation problem with improved harmony search

控制与决策. 2020, 35(11): 2743–2751 <https://doi.org/10.13195/j.kzyjc.2019.0242>

### 基于强化学习的小型无人直升机有限时间收敛控制设计

Finite time control based on reinforcement learning for a small-size unmanned helicopter

控制与决策. 2020, 35(11): 2646–2652 <https://doi.org/10.13195/j.kzyjc.2019.0328>

# 阴影条件下基于迁移强化学习的光伏系统最大功率跟踪

杨 博<sup>1</sup>, THIDAR Swe<sup>1</sup>, 钟林恩<sup>1</sup>, 束洪春<sup>1</sup>, 张孝顺<sup>2†</sup>, 余 涛<sup>3</sup>

(1. 昆明理工大学 电力工程学院, 昆明 650500; 2. 汕头大学 工学院,  
广东 汕头 515063; 3. 华南理工大学 电力学院, 广州 510640)

**摘 要:** 在光伏系统中, 光伏阵列往往会受到阴影条件 (partial shading condition, PSC) 的影响, 造成光伏系统输出功率偏低以及功率-电压 ( $P$ - $V$ ) 特性曲线出现多峰值的现象, 从而导致常规最大功率跟踪 (maximum power point tracking, MPPT) 算法易陷入局部最优的问题. 对此, 设计一种基于迁移强化学习 (transfer reinforcement learning, TRL) 的 MPPT 算法. 该算法将连续变量的动作空间分解为若干个小范围的子搜索空间, 从而有效提高 TRL 的学习效率. 同时, 引入知识迁移, 即将旧任务的最优知识矩阵应用到新任务中, 进而大幅提高 TRL 的收敛速度. 通过对 3 种算例的研究, 即恒温变光照强度、变温变光照强度和香港实地测试, 其仿真结果表明, 与传统增量电导法 (incremental conductance, INC)、遗传算法 (genetic algorithm, GA)、粒子群优化 (particle swarm optimization, PSO) 算法、人工蜂群 (artificial bee colony, ABC) 算法、布谷鸟算法 (cuckoo search algorithm, CSA)、教-学优化 (teaching-learning based optimization, TLBO) 算法以及 Q 学习算法相比, TRL 能在 PSC 下实现最快速的全局最大功率跟踪, 同时具有最小的功率波动. 最后, 基于 dSpace 的硬件在环实验验证了 TRL 的硬件可行性.

**关键词:** 光伏系统; 最大功率跟踪; 迁移强化学习; 动作空间分解

中图分类号: TM7

文献标志码: A

DOI: 10.13195/j.kzyjc.2019.0412

开放科学(资源服务)标识码(OSID):



**引用格式:** 杨博, THIDAR Swe, 钟林恩, 等. 阴影条件下基于迁移强化学习的光伏系统最大功率跟踪[J]. 控制与决策, 2020, 35 (12): 2939-2949.

## Transfer reinforcement learning based maximum power point tracker of PV systems under partial shading condition

YANG Bo<sup>1</sup>, THIDAR Swe<sup>1</sup>, ZHONG Lin-en<sup>1</sup>, SHU Hong-chun<sup>1</sup>, ZHANG Xiao-shun<sup>2†</sup>, YU Tao<sup>3</sup>

(1. Faculty of Electric Power Engineering, Kunming University of Science and Technology, Kunming 650500, China; 2. School of Engineering, Shantou University, Shantou 515063, China; 3. College of Electric Power, South China University of Technology, Guangzhou 510640, China)

**Abstract:** In photovoltaic (PV) systems, PV arrays are often affected by shadows which results in a reduction of generated power. The power-voltage ( $P$ - $V$ ) characteristics of PV array will contain multiple peaks under partial shading condition (PSC), while conventional maximum power tracking (MPPT) techniques are easy to fall into a local optimum. Therefore, this paper designs a novel MPPT algorithm based on transfer reinforcement learning (TRL). The algorithm decomposes the original large-scale search space into several small-scale sub-search spaces, which can effectively improve the global search ability of TRL. Meanwhile, the knowledge transfer is adopted to transfer the optimal knowledge matrix of the old task to the new task, such that the convergence rate of TRL could be improved significantly. Three cases are carried out, e.g., constant temperature and varying solar irradiation, varying temperature and varying solar irradiation, as well as HongKong field test. Simulation results demonstrate that TRL can achieve the fastest MPPT under PSC and lowest power fluctuations in comparison to incremental conductance (INC), genetic algorithm (GA), particle swarm optimization (PSO), artificial bee colony (ABC), cuckoo search algorithm (CSA), teaching-learning based optimization (TLBO), and Q-learning. Finally, dSpace based hardware-in-loop (HIL) test verifies the implementation feasibility of TRL.

**Keywords:** photovoltaic system; maximum power point tracking; transfer reinforcement learning; space decomposition

收稿日期: 2019-04-05; 修回日期: 2019-07-13.

基金项目: 国家自然科学基金项目(61963020); 云南省应用基础研究计划项目-青年项目(2018FD036).

责任编委: 孙秋野.

<sup>†</sup>通讯作者. E-mail: xszhang1990@sina.cn.

## 0 引言

光伏系统在实际运行过程中,由于周围建筑物、树木和云层的遮挡,即阴影条件(partial shading condition, PSC),使得光伏阵列表面受到不均匀的光照强度,极易对光伏阵列运行产生不利影响<sup>[1]</sup>.为改善这一情况,通常加入旁路二极管以提高光伏系统的输出功率,这使得整个输出伏安特性曲线呈阶梯状.而传统的最大功率跟踪算法仅能在均匀光照强度下收敛至最大功率点,例如扰动观测法(perturb & observe, P&O)<sup>[2]</sup>.然而,上述算法在最大功率点(maximum power point, MPP)附近通常会产生持续的功率波动,同时在PSC下易陷入局部最大功率点(local maximum power point, LMPP),从而导致发电效率较低等问题.

近年来,国内外学者提出了各种先进的MPPT算法来解决PSC下的全局最大功率点(global maximum power point, GMPP)跟踪问题,包括遗传算法(genetic algorithm, GA)<sup>[3]</sup>、粒子群算法(particle swarm optimization, PSO)<sup>[4]</sup>、人工蜂群算法(artificial bee colony, ABC)<sup>[5]</sup>、布谷鸟搜索算法(cuckoo search algorithm, CSA)<sup>[6]</sup>、教学优化算法(teaching-learning based optimization, TLBO)<sup>[7]</sup>等启发式智能优化算法,然而这些启发式算法存在以下两个问题<sup>[8]</sup>:

1) 收敛不稳定性.与确定性优化算法不同,启发式算法由于其随机搜索机制,即使在相同运行条件下得到的最优解也可能存在较大差异.因此,在实际运用中,启发式算法易导致光伏系统产生较大的功率波动.

2) 难以平衡最优解与收敛时间的关系.为了获得高质量的最优解,启发式算法通常具有较大的种群规模和较多的迭代次数,故需要消耗较长的计算时间.为满足光伏系统的快速MPPT要求,往往不可避免地减少种群规模和迭代次数,这将导致解的质量降低.

随着深度学习技术的成熟,人工智能正在从尖端技术逐步普及于各行各业.其中,强化学习(reinforcement learning, RL)是一种以环境反馈作为输入且适应不确定环境的机械学习方法<sup>[9]</sup>,作为一种“奖惩”式的评价学习算法,它不依赖于系统的数学模型,仅利用来自环境的评价信号更新自身参数.然而,与传统机器学习算法一样,RL在学习新的任务时,智能体需要新的环境中进行随机探索与试错以获得新的知识<sup>[10]</sup>.如果新任务与历史任务(源任务)相关性较高,则智能体探索的动作就会出现较大重复性,耗费

较多的计算时间<sup>[11]</sup>.为避免智能体在新环境中的盲目探索,学者们开始将迁移学习引入强化学习算法领域,尝试利用从历史任务的已学知识中提炼出有效信息来指导或加速新任务的学习,形成了一类全新的迁移强化学习(transfer reinforcement learning, TRL)算法.

一般来说,TRL可以划分为行为迁移和信息迁移两种类型<sup>[12]</sup>.行为迁移通常利用已学到的策略或子过程来指导智能体对新任务的学习,所以侧重源任务与新任务之间的相关性计算.信息迁移则把源任务当成新任务的监督信息来改善其学习的性能,例如关系强化学习<sup>[13]</sup>.目前,应用到电力系统优化与控制领域的TRL主要属于行为迁移类型.文献[14]结合蜂群寻优机制,提出了迁移蜂群优化算法,并成功应用到无功优化问题中.文献[15]则在迁移强化学习中引入细菌觅食的搜索机制,实现了电力系统风险调度的快速求解.文献[16]将迁移强化学习扩展成多智能体形式,解决了自动发电控制的功率动态分配分散式优化问题.在这些研究中,通过对具体问题知识迁移的结合,迁移强化学习算法在保证求解质量的同时,寻优速度可明显提升,最快可达到其他传统启发式算法速度的100倍以上.因此,TRL可以快速找到求解对象的较高质量最优解,这很符合光伏系统的快速最大功率控制需求.然而,上述TRL方法<sup>[14-16]</sup>主要是基于离散控制变量设计的,无法满足连续变量的高效寻优需求.

为进一步提升TRL的寻优性能,本文引入动作空间分解策略将连续变量的搜索空间分解为多个子搜索空间,以提高TRL的全局搜索能力.基于Matlab仿真实验对TRL与GA、PSO、ABC、CSA、TLBO和Q学习算法进行对比.最后,通过基于dSpace的硬件在环实验(hardware-in-loop, HIL)验证了TRL的硬件可行性.

## 1 阴影条件

光伏阵列在运行中经常被树木或云层遮挡导致其所受到的光照强度不均匀而产生局部阴影的现象,从而导致光伏系统输出功率降低,影响光伏电池的整体寿命<sup>[17]</sup>.光伏阵列组成的回路中通常需配置一个或几个旁路二极管以消除功率失配时造成的“热斑效应”.此外,在每一条光伏阵列串末端需串联防逆流二极管,以防止不同光伏阵列串间电流倒流.图1(a)和图1(b)分别给出了均匀光照强度和PSC下光伏系统运行图.

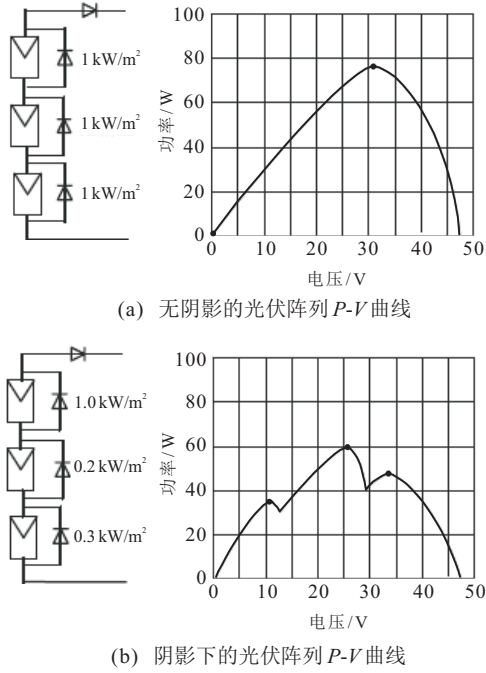


图1 光伏阵列不同工况下的运行图及特性曲线

如图1(b)所示,在阴影条件下,随着电压的变化,光伏系统会产生多个功率峰值.对于传统方法(例如P&O)而言,其最终收敛的功率点完全取决于初始点及其参数的选择,如果初始功率点靠近某个峰值点,计算步长又设置得比较小,则容易陷入某个局部峰值点,无法逼近全局最大功率点.

## 2 基于空间分解的迁移强化学习

### 2.1 动作空间分解

传统Q学习算法只能实现离散动作集的寻优<sup>[18]</sup>,而本文的光伏MPPT应用对象控制变量是连续的,所以本文将算法原始动作空间分解成多个小范围的搜索动作子空间,如图2所示.

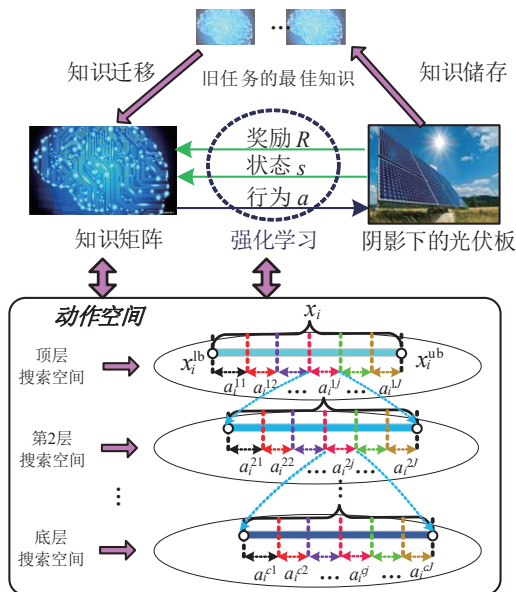


图2 空间分解的TRL原理

设搜索空间层数为 $c$ ,每一层的离散动作数量为 $J$ ,即可将每个控制变量的寻优范围转化为由 $(c \times J)$ 个组合动作构成的动作空间.因此,被控变量 $x_i$ 的优化精度可计算如下:

$$OA_i = \frac{x_i^{ub} - x_i^{lb}}{c \cdot J}, \quad (1)$$

其中 $x_i^{lb}$ 和 $x_i^{ub}$ 分别是第 $i$ 个被控变量的下限和上限.

从式(1)可以看出,搜索空间层数 $c$ 和动作个数 $J$ 设置得越大,优化精度越高,但计算速度会有所下降.因此,具体参数的设置需要根据具体优化问题的复杂度及时间要求来定.对于光伏系统的MPPT而言,其时间要求更高,因此,参数 $c$ 和 $J$ 不宜设置较大.若每层中的动作次数设置为10( $J = 10$ ),则当 $c = 6$ 时,0到1之间的连续控制变量可实现相同的精度( $10^{-6}$ ).这意味着所选动作次数可以从 $10^6$ 减少到10.因此,空间分解可显著提高Q学习的学习速率和控制精度.

在选择所有层中的所有动作之后,可以确定可控变量的解为

$$x_i = x_i^{c,lb} + a_i^{c,j} \cdot (x_i^{c,ub} - x_i^{c,lb})/J. \quad (2)$$

$$x_i^{l,lb} =$$

$$\begin{cases} x_i^{lb}, l = 1; \\ x_i^{l-1,lb} + a_i^{l-1,j} \cdot \frac{(x_i^{l-1,ub} - x_i^{l-1,lb})}{J}, \text{ otherwise.} \end{cases} \quad (3)$$

$$x_i^{l,ub} =$$

$$\begin{cases} x_i^{ub}, l = 1; \\ x_i^{l-1,ub} + a_i^{l-1,j} \cdot \frac{(x_i^{l-1,ub} - x_i^{l-1,lb})}{J}, \text{ otherwise.} \end{cases} \quad (4)$$

其中: $x_i^{l,lb}$ 和 $x_i^{l,ub}$ 分别是第 $l$ 层搜索空间的下界和上界; $a_i^{l,j}$ 是第 $l$ 层搜索空间中的第 $j$ 个动作.

### 2.2 知识更新

根据Q学习的学习机理,基于状态动作的反馈奖励来更新知识矩阵.通过组合空间分解,每个搜索空间层的知识矩阵更新<sup>[19]</sup>如下:

$$\begin{aligned} Q_{i,k}^l(s_{i,k}^l, a_{i,k}^l) = & Q_{i,k}^l(s_{i,k}^l, a_{i,k}^l) + \alpha [R_{i,k}^l(s_{i,k}^l, a_{i,k}^l, s_{i,k+1}^l, a_{i,k}^l) + \\ & \gamma \max_{a \in A_i^l} Q_{i,k}^l(s_{i,k+1}^l, a) - Q_{i,k}^l(s_{i,k}^l, a_{i,k}^l)]. \end{aligned} \quad (5)$$

其中: $Q_i^l$ 表示第 $i$ 个被控变量的第 $l$ 层搜索空间的知识矩阵; $(s_{i,k}^l, a_{i,k}^l)$ 是在第 $k$ 次迭代时执行的状态动作对( $k = 1, 2, \dots, k_{\max}$ ), $k_{\max}$ 是最大迭代步数; $\alpha$ 是学习因子; $\gamma$ 是折扣因子; $R_i^l$ 是奖励动作; $A_i^l$ 表示第 $l$

层搜索空间的动作空间。

由于传统Q学习在动态环境中使用单个agent进行搜索,每次迭代只能更新每个知识矩阵的一个元素.这不可避免地导致学习速度较慢,进而难以在光伏系统实时控制中快速地获得高质量的最优解.所以,采用协作机制进一步加速学习速率,更新TRL的每个知识矩阵<sup>[18]</sup>,即

$$\begin{aligned} Q_{i,k}^l(s_{i,k}^{l,m}, a_{i,k}^{l,m}) = & Q_{i,k}^l(s_{i,k}^{l,m}, a_{i,k}^{l,m}) + \alpha [R_{i,k}^{l,m}(s_{i,k}^{l,m}, s_{i,k+1}^{l,m}, a_{i,k}^{l,m}) + \\ & \gamma \max_{a \in A_i^l} Q_{i,k}^l(s_{i,k+1}^{l,m}, a) - Q_{i,k}^l(s_{i,k}^{l,m}, a_{i,k}^{l,m})], \end{aligned} \quad (6)$$

$$m = 1, 2, \dots, M,$$

其中  $M$  表示协作机制中的种群规模。

### 2.3 全局搜索和局部探索

强化学习算法在学习过程中需要根据已有的知识矩阵在不同状态下选择动作:如果都采用贪婪动作,则学习速度加快,但容易收敛到局部最优解;反之,如果选择动作比较随机,则全局搜索能力更强,但学习效率较低<sup>[20]</sup>.为平衡上述矛盾,本文引入  $\varepsilon$ -贪婪策略( $\varepsilon$ -greedy rule)选择动作<sup>[21]</sup>,即算法可通过设置合适的参数  $\varepsilon$  来平衡全局搜索(随机动作选择)和局部搜索(贪婪动作选择),既以较大的概率选择贪婪动作,又保留一定的概率选择随机动作,从而达到平衡这两者的目的.具体的动作可选择如下:

$$a_{i,k+1}^{l,m} = \begin{cases} \arg \max_{a_i^l \in A_i^l} Q_{i,k}^l(s_{i,k+1}^{l,m}, a_i^l), & q_0 < \varepsilon; \\ a_{\text{rand}}, & \text{otherwise.} \end{cases} \quad (7)$$

其中:  $q_0$  是 0 到 1 之间的均匀随机数,  $\varepsilon$  是贪婪概率,  $a_{\text{rand}}$  表示动作空间中的随机动作。

### 2.4 知识迁移

式(6)给出的知识矩阵相当于算法在寻优过程中形成的搜索知识,可为智能体提供有效的搜索引导.如果不采用迁移,则强化学习因缺乏先验知识,在初始学习阶段会选择比较盲目的探索,经过一系列与环境交互训练之后,才能收敛到当前新任务的最优解,这样就会耗费较长的学习训练时间.与之相比,本文在引入知识迁移后,可为强化学习提供有效的先验知识,即在初始学习阶段可以实施高效的寻优,避免盲目的探索,从而明显减少训练时间.换言之,在光伏MPPT的较短控制时间周期内,引入知识迁移后的强化学习找到更高质量最优解的概率会更大.本文采用的知识迁移<sup>[22]</sup>是指,从历史任务的最优知识矩阵中提炼出近似新任务的最优知识矩阵.本文选取与新

任务相似度最高的旧任务进行知识迁移,即

$$Q_i^{n0} = r \cdot Q_i^{s*} + (1 - r) \cdot Q_i^{\text{initial}}. \quad (8)$$

其中:  $Q_i^{n0}$  是新任务的第  $i$  个被控变量的近似最优知识矩阵;  $Q_i^{s*}$  是最相似的旧任务的第  $i$  个被控变量的最优知识矩阵;  $Q_i^{\text{initial}}$  是没有知识迁移的初始知识矩阵;  $r$  表示最相似的旧任务与新任务之间的相似度,  $0 \leq r \leq 1$ .

## 3 光伏系统最大功率跟踪的TRL设计

### 3.1 被控变量和动作空间

为了获得光伏系统的GMPP,选择输出电压  $V_{\text{pv}}$  作为被控变量,其中整个搜索空间被分解为4层.在每一层中,搜索空间在相应范围内被均匀地离散化为10个动作。

### 3.2 奖励函数

对于已知的输出电压  $V_{\text{pv}}$ ,光伏系统可在给定的光照强度、温度和阴影条件下产生相应的功率.在TRL中,具有更高质量解的个体将会获得更大的奖励.基于此机制,奖励函数可设计如下:

$$R_{i,k}^{l,m}(s_{i,k}^{l,m}, s_{i,k+1}^{l,m}, a_{i,k}^{l,m}) = \begin{cases} \max_{m=1,2,\dots,M} f(V_{\text{pv}}^m), & (s_{i,k}^{l,m}, a_{i,k}^{l,m}) \in \text{SA}_k^{\text{best}}; \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

其中:  $V_{\text{pv}}^m$  是第  $m$  个体获得的解;  $\text{SA}_k^{\text{best}}$  表示在第  $k$  次迭代中具有最大功率输出的最佳个体的搜索动作状态。

### 3.3 不同气候条件下的知识迁移

光伏系统的发电功率主要是由光照强度、温度和阴影情况3个因素决定的,因此,不同时刻的MPPT控制任务之间的相似度也可以用这3个因素来评价.另一方面,天气在短时间内不会发生明显变化,因此,两个相邻优化任务之间的相似度通常是非常高的.如图3所示,选择最相邻旧任务的最佳知识矩阵用于新任务的知识迁移,式(8)中描述的相似度可设计为

$$r = 1 - \frac{|T_c^n - T_c^p|}{T_{\text{ref}}} - \prod_{w=1}^{N_s \cdot N_p} \frac{|S_w^n - S_w^p|}{S_{\text{ref}}}. \quad (10)$$

其中:  $T_c^n$  和  $T_c^p$  分别代表新旧任务中的温度,  $S_w^n$  和  $S_w^p$  分别是新旧任务中第  $w$  个光伏电池的光照强度。

如式(10)所示,相似度与上述3个因素直接相关,如果不同任务之间的气象条件差异较小,则说明它们之间的相似度较高,最高时为1(即任务之间气象条件完全一样)。

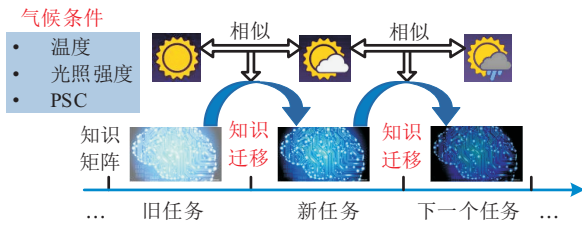


图3 相邻两个任务之间的TRL知识迁移

### 3.4 参数设置

从式(1)~(10)可以看出,TRL涉及到7个主要参数,这些参数的选择将影响算法在光伏MPPT的应用效果。通常,这些参数可按以下原则进行选取:

1) 动作空间层数 $c$  ( $c \geq 1$ ):  $c$ 越大,表明寻优的精度越高,但搜索时间更长。对于光伏MPPT而言, $c$ 应设置为较小的值,以满足在线优化需求。

2) 每一层离散动作数量 $J$  ( $J > 1$ ): 与参数 $c$ 选取原则一样,不宜设置过大。

3) 学习因子 $\alpha$  ( $0 < \alpha < 1$ ):  $\alpha$ 越大,算法越快收敛,但更容易陷入局部最优解;反之,算法收敛速度越慢,但全局搜索能力更强。在引入知识迁移之后,算法已具有较好的先验知识,因此 $\alpha$ 可以设置为较小的值。

4) 折扣因子 $\gamma$  ( $0 < \gamma < 1$ ): 表征知识矩阵在更新过程中对过去奖励值的折扣。如果长期奖励累积值对求解问题影响更大,则 $\gamma$ 应取较大的值;反之亦然。对于光伏MPPT而言,立即奖励值对控制效果影响更大,因此, $\gamma$ 应设置为较小的值。

5) 贪婪概率 $\varepsilon$  ( $0 < \varepsilon < 1$ ):  $\varepsilon$ 越大,代表算法选择贪婪动作的概率越大,算法能更快收敛,但也更容易陷入局部最优解;反之, $\varepsilon$ 越小,算法学习速度越慢,但全局搜索能力更强。对于光伏MPPT而言, $\varepsilon$ 同样需要设置为较大的值,以满足在线优化需求。

6) 最大迭代步数 $k_{\max}$  ( $k_{\max} > 1$ ):  $k_{\max}$ 越大,代表算法获得更高质量最优解的概率越大,但耗费的计算时间越长;反之, $k_{\max}$ 越小,虽然可以减少计算时间,但算法更容易陷入局部最优。在引入知识迁移之后,算法已具有较好的先验知识,因此, $k_{\max}$ 可以设置为较小的值。

7) 种群规模 $M$  ( $M \geq 1$ ): 与参数 $k_{\max}$ 选取原则一样,在引入知识迁移之后, $M$ 可以设置为较小的值。通常,在不同环境下光伏发电系统的运行工况将发生变化,需要设置不同的最优参数以满足不同环境下的MPPT最优控制。然而,这将耗费较大的仿真测试工作量,也会增加算法在线优化控制的复杂性。因此,本文根据参数选取原则,结合仿真测试结果确定

出一组固定的最优参数(见表1),以满足光伏系统在不同环境下的最优MPPT控制需求。

表1 TRL主要参数

| 参数            | 范围                    | 取值      |
|---------------|-----------------------|---------|
| $J$           | $J > 1$               | 10      |
| $c$           | $c > 1$               | 4       |
| $\alpha$      | $0 < \alpha < 1$      | 0.01    |
| $\gamma$      | $0 < \gamma < 1$      | 0.000 1 |
| $\varepsilon$ | $0 < \varepsilon < 1$ | 0.9     |
| $k_{\max}$    | $k_{\max} > 1$        | 5       |
| $M$           | $M > 1$               | 5       |

需要说明的是,如果算法不引入迁移学习,则参数 $\alpha$ 、 $\gamma$ 、 $k_{\max}$ 和 $M$ 需重新设置,其中参数 $k_{\max}$ 和 $M$ 需要适当增加,才能获得与引入迁移学习的TRL算法接近的优化性能,这也将增加算法的计算时间。对于本文构建的光伏系统MPPT而言,经过仿真测试,如果不引入迁移学习,则这两个参数需分别设置为50和30。因此,与不引入迁移学习的算法相比,TRL计算时间大概可缩短为原来的1/60。

### 3.5 总体流程

光伏系统在PSC下实现MPPT的TRL总体流程如图4所示。首先,输出电压的原始搜索空间被分解为4层较小的子搜索空间;其次,根据相似的气候条

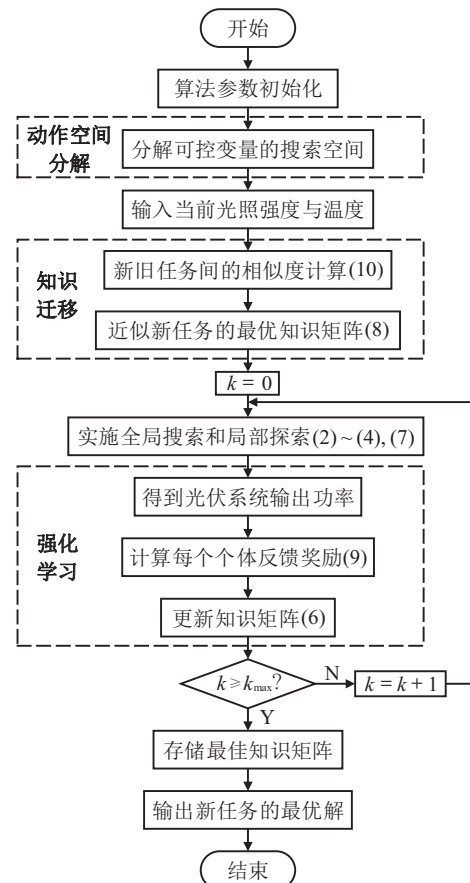


图4 实现MPPT的TRL流程

件实现新旧任务之间的知识迁移;再次,TRL可在迭代中通过持续搜索来更新自身的知识矩阵;最后,输出最优解,即最优输出电压,从而实现在PSC下的MPPT.

#### 4 算例分析

在本节中将分析3个算例,即恒温变光照强度、变温变光照强度和香港实地测试,分别与INC<sup>[2]</sup>、GA<sup>[3]</sup>、PSO<sup>[4]</sup>、ABC<sup>[5]</sup>、CSA<sup>[6]</sup>、TLBO<sup>[7]</sup>、Q学习<sup>[18]</sup>进行对比,以评估PSC下TRL的MPPT性能.按照表1给出的参数,算法将可控变量占空比的区间 $[0, 1]$ 划分为4层搜索空间,每一层有10个区间范围.光伏系统参数的选取见文献[8].另外,分别设置额定光照强度和温度为 $1000 \text{ W/m}^2$ 和 $25^\circ\text{C}$ ,进而有 $P_{\text{pv}} = 51.716 \text{ W}$ ,  $V_{\text{dc}} = 18.47 \text{ V}$ ,  $I_{\text{pv}} = 2.8 \text{ A}$ .仿真在Matlab/Simulink 2016a上进行,采用ode45 (Dormand-Prince)变步长求解器.

##### 4.1 参数灵敏度分析

如表1所示,TRL有7个主要参数会对其寻优性能产生影响.本节选择其中两个主要参数(学习因子 $\alpha$ 和贪婪概率 $\varepsilon$ )对光伏能量输出的灵敏度进行分析,分别将它们取值为 $\{0.1, 0.2, \dots, 0.9\}$ ,参数组合数量为81,如图5所示.

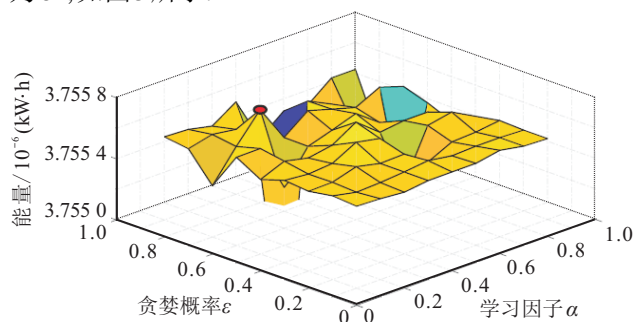


图5 算法参数对光伏能量的影响

由图5可以发现:1)贪婪概率取值较大时,光伏系统输出能量总体较大;2)学习因子取值大小对光伏系统输出能量影响较为随机,但取值较小时,能在某一参数组合下获得最高的能量输出.

##### 4.2 恒温变光照强度

为研究TRL在恒温变光照强度下的MPPT性能,在光伏阵列上施加4个连续的光照强度阶跃信号,如图6所示,温度保持在额定值 $25^\circ\text{C}$ .

另外,图7提供了恒温变光照强度下不同算法的结果.由图7可见,当光照强度突变时,INC陷入LMPP,启发式算法收敛至GMPP.显然,TRL的功率波动最小,尤其是,TRL所产生的能量比INC的能量高24.64%,这表明知识迁移能够有效地提高TRL的

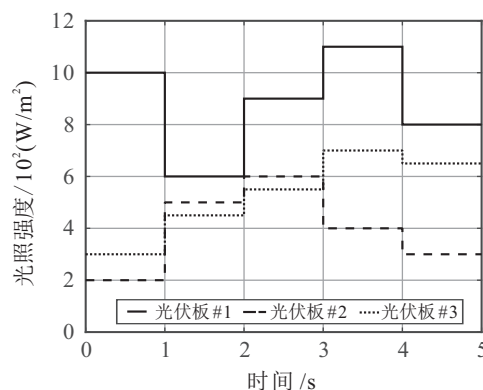
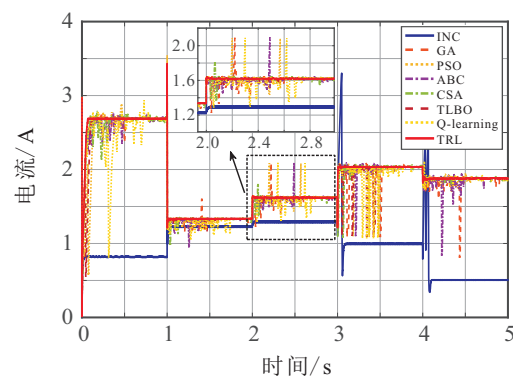
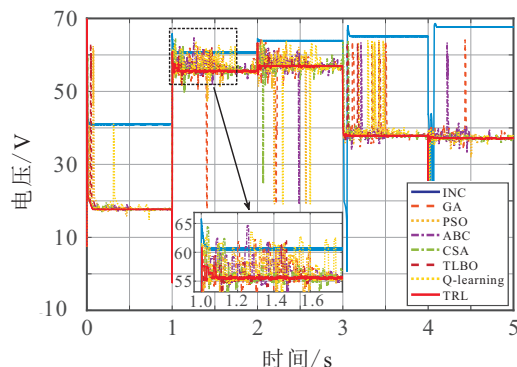


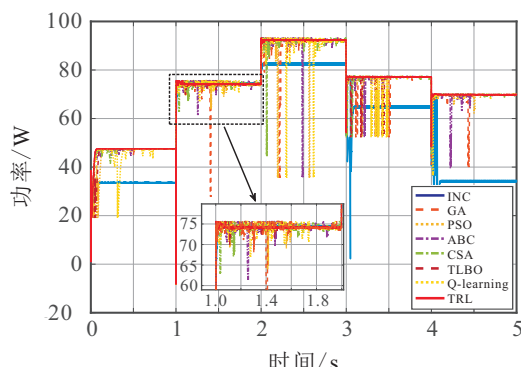
图6 光照强度



(a) 电流波形



(b) 电压波形



(c) 功率波形

图7 恒温变光照强度不同算法光伏系统响应曲线

最优解搜索质量.其中,在第1个光照强度阶跃信号下,TRL最终收敛的最优解为0.8752(DC变换器的占空比),即表示算法选择的动作组合为:第1层搜索空间的第9个动作(对应区间 $(0.8, 0.9]$ )、第2层搜索空

间的第8个动作(对应区间 $(0.87, 0.88]$ )、第3层搜索空间的第6个动作(对应区间 $(0.875, 0.876]$ )、第4层搜索空间的第3个动作(对应区间端点 $0.8752$ )。

#### 4.3 变温变光照强度

为研究TRL在变温变光照强度下的MPPT性能,在光伏阵列上施加4个连续的光照强度和温度变化信号,如图8所示。另外,图9显示了在光照强度和温度逐渐变化下8种MPPT算法的结果。可以看出,除TRL之外的其他算法在光照强度和温度的逐渐变化中易产生较大的功率波动。另一方面,也可发现TRL在线寻优效果明显优于Q学习算法,这也说明了动作空间分解和知识迁移能有效提升强化学习的MPPT应用效果。由于知识迁移避免了盲目的初始寻优,TRL能显著地降低光伏系统的功率波动。同样,TRL仍然在所有算法中获得光伏系统的最高的能量输出,比INC获得的能量高11.65%。

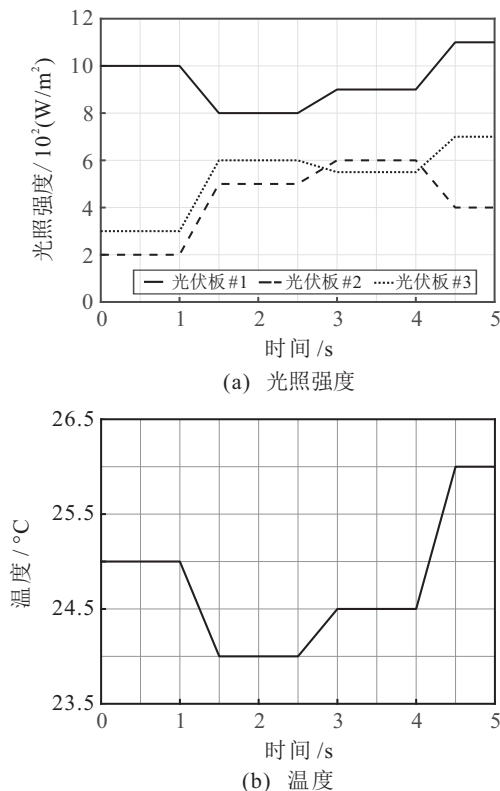


图8 光照强度与温度曲线

#### 4.4 香港实地测试

本文采用实际测量的光照强度和温度数据,进一步评估TRL的MPPT性能。测量数据选自2017年香港四季中的4个典型日,实际光照强度和温度曲线如图10所示,数据测量间隔为10 min。测量装置安装于北纬 $22.3^\circ$ ,东经 $114.2^\circ$ 。为模拟真实的PSC环境,光伏电池#1至光伏电池#3的光照强度分别设置为实际光照强度的100%、80%和50%。

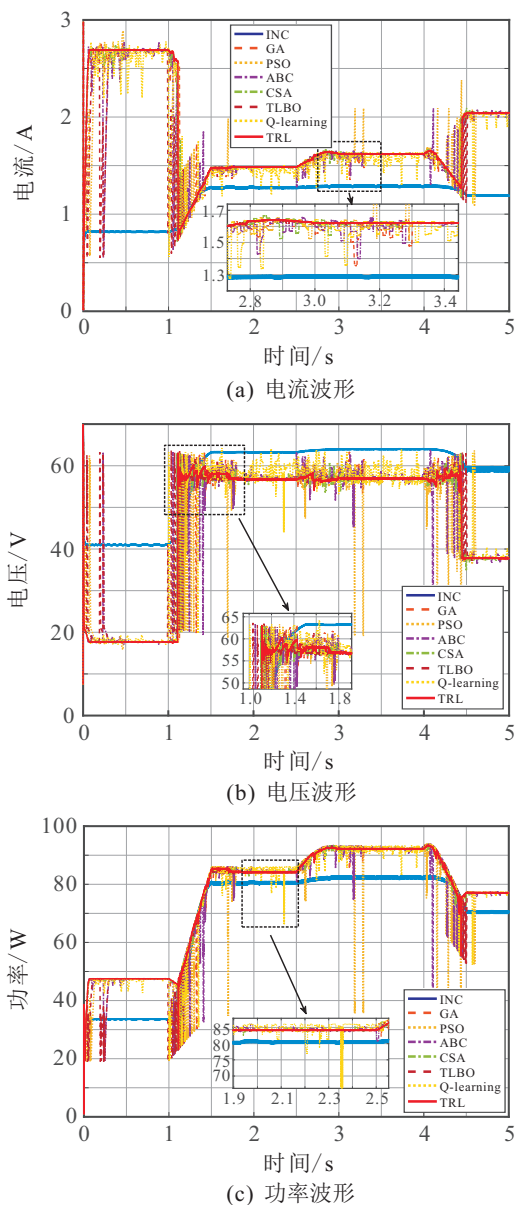


图9 变温变光照强度不同算法光伏系统响应曲线

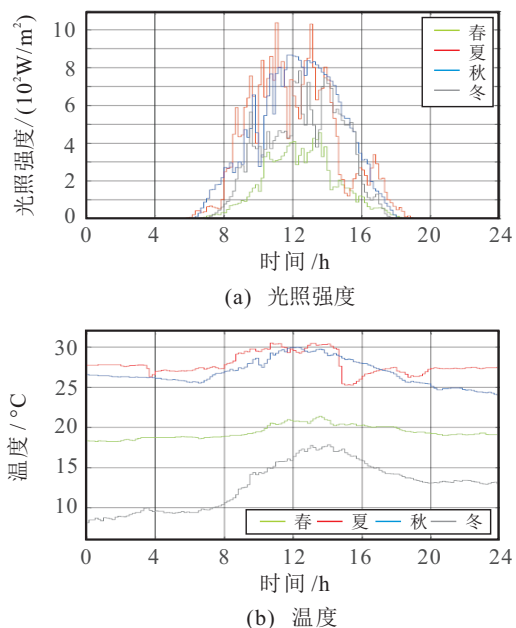


图10 香港实际光照强度和温度曲线

图11显示了香港四季中不同算法MPPT的仿真曲线. 由图11可知,TRL和启发式算法都能获得比INC更多的能量,其中TRL在春季获得的能量是INC

的115.52%. 同时,在所有季节中,TRL的表现性能均优于启发式算法. 如TRL在春季获得的能量是INC的103.32%.

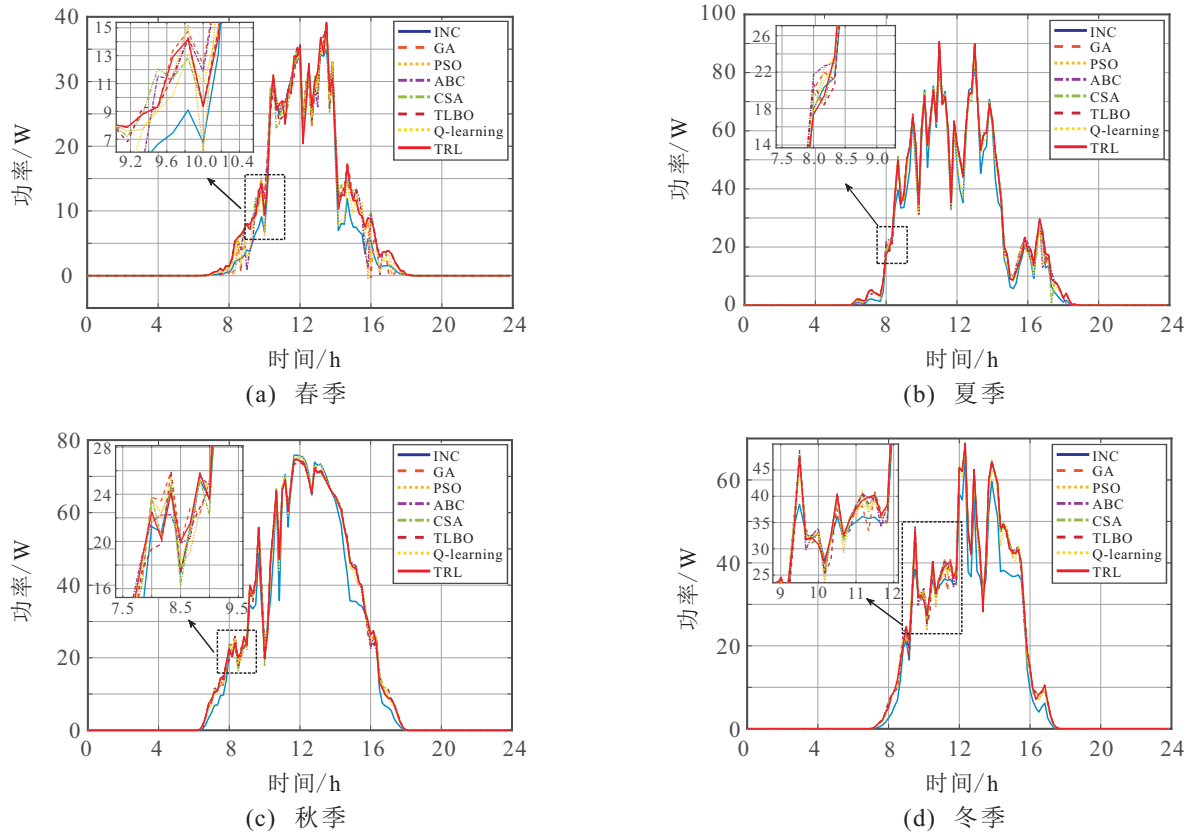


图 11 香港四季典型日中8种算法光伏系统响应曲线

表2 3种算例下各算法统计指标

| 算例            | 指标                                    | INC             | GA        | PSO             | ABC             | CSA             | TLBO          | Q-learning | TRL              |
|---------------|---------------------------------------|-----------------|-----------|-----------------|-----------------|-----------------|---------------|------------|------------------|
| 恒温变光照强度       | 能量/ $(10^{-6}\text{kW}\cdot\text{h})$ | 80.258 5        | 99.269 5  | 99.192 4        | 99.550 3        | 99.581 2        | 99.573 8      | 98.876 8   | <b>100.033 3</b> |
|               | $\Delta v^{\max}/\%$                  | 43.596 7        | 34.319 4  | 34.312 2        | 34.179 1        | <b>32.081 6</b> | 34.192 7      | 34.427 9   | 33.978 1         |
|               | $\Delta v^{\text{avg}}/\%$            | 0.032 4         | 0.007 8   | 0.008 0         | 0.007 5         | 0.007 6         | 0.007 5       | 0.008 6    | <b>0.006 6</b>   |
| 变温变光照强度       | 能量/ $(10^{-6}\text{kW}\cdot\text{h})$ | 94.570 1        | 105.323 3 | 103.694 9       | 104.487 2       | 105.374 7       | 104.271 1     | 104.897 0  | <b>105.588 0</b> |
|               | $\Delta v^{\max}/\%$                  | <b>24.114 1</b> | 21.652 1  | 21.992 1        | 21.825 4        | 21.641 6        | 21.870 6      | 21.740 1   | 21.597 8         |
|               | $\Delta v^{\text{avg}}/\%$            | 0.030 0         | 0.007 4   | 0.009 9         | 0.009 2         | 0.007 3         | 0.008 5       | 0.008 2    | <b>0.006 7</b>   |
| 香港实测<br>春季典型日 | 能量/ $(\text{kW}\cdot\text{h})$        | 0.140 3         | 0.156 9   | 0.159 3         | 0.158 6         | 0.159 2         | 0.158 4       | 0.150 3    | <b>0.160 5</b>   |
|               | $\Delta v^{\max}/\%$                  | 257.46          | 183.48    | 178.24          | <b>163.03</b>   | 212.33          | 216.22        | 222.13     | 227.40           |
|               | $\Delta v^{\text{avg}}/\%$            | <b>17.664 5</b> | 19.307 1  | 20.296 4        | 21.242 1        | 18.319 2        | 18.789 4      | 21.539 2   | 18.819 3         |
| 香港实测<br>夏季典型日 | 能量/ $(\text{kW}\cdot\text{h})$        | 0.390 2         | 0.428 2   | 0.428 8         | 0.428 5         | 0.426 8         | 0.427 2       | 0.409 7    | <b>0.429 0</b>   |
|               | $\Delta v^{\max}/\%$                  | 174.36          | 166.10    | 158.78          | 166.02          | 157.23          | 160.78        | 174.20     | <b>156.80</b>    |
|               | $\Delta v^{\text{avg}}/\%$            | 25.932 9        | 22.785 7  | <b>22.403 7</b> | 23.229 6        | 23.298 4        | 22.836 7      | 24.169 2   | 22.774 8         |
| 香港实测<br>秋季典型日 | 能量/ $(\text{kW}\cdot\text{h})$        | 0.422 2         | 0.449 5   | 0.448 6         | 0.449 0         | 0.449 5         | 0.448 7       | 0.448 3    | <b>0.448 9</b>   |
|               | $\Delta v^{\max}/\%$                  | 173.21          | 106.35    | 113.15          | 105.17          | 112.72          | <b>105.04</b> | 124.74     | 107.12           |
|               | $\Delta v^{\text{avg}}/\%$            | 13.341 2        | 11.458 3  | 11.833 1        | <b>11.247 7</b> | 12.276 5        | 11.770 8      | 12.117 8   | 11.838 7         |
| 香港实测<br>冬季典型日 | 能量/ $(\text{kW}\cdot\text{h})$        | 0.281 2         | 0.320 3   | 0.319 2         | 0.319 6         | 0.320 4         | 0.318 0       | 0.315 3    | <b>0.321 4</b>   |
|               | $\Delta v^{\max}/\%$                  | 234.77          | 212.16    | <b>188.74</b>   | 183.83          | 198.02          | 204.26        | 188.97     | 183.75           |
|               | $\Delta v^{\text{avg}}/\%$            | <b>16.457 7</b> | 17.607 7  | 18.365 3        | 18.070 7        | 17.298 0        | 18.039 5      | 17.477 1   | 17.367 2         |
| 平均计算时间/ms     |                                       | <b>0.01</b>     | 7.54      | 2.16            | 1.39            | 3.29            | 4.04          | 3.67       | 3.59             |

#### 4.5 统计分析

为定量评估光伏系统的功率振荡幅度,引入如下两个指标:

$$\Delta v^{\text{avg}} = \frac{1}{T-1} \sum_{t=2}^T \frac{|P_{\text{out}}(t) - P_{\text{out}}(t-1)|}{P_{\text{out}}^{\text{avg}}}, \quad (11)$$

$$\Delta v^{\text{max}} = \max_{t=2,3,\dots,T} \frac{|p_{\text{out}}(t) - P_{\text{out}}(t-1)|}{P_{\text{out}}}. \quad (12)$$

其中:  $\Delta v^{\text{avg}}$  和  $\Delta v^{\text{max}}$  分别表示光伏系统输出功率的平均振荡指标和最大振荡指标,  $t$  是时间,  $T$  是总运行时间,  $P_{\text{out}}^{\text{avg}}$  是在总迭代次数内光伏系统输出功率的平均值。

表2给出了3种算例下8种算法的统计结果. 可见TRL在3种算例下均获得最大能量并具有最小的功率振荡(最优值加粗标记).

### 5 硬件在环实验

本节基于 dSpace 进行 HIL 实验以验证 TRL 的硬件可行性, HIL 实验平台如图12所示. TRL 算法(1)~(10)置于 DS1104 平台, 其采样频率  $f_c = 10 \text{ kHz}$ ; 光伏系统温度和光照强度置于 DS1006 平台, 其采样频率  $f_s = 100 \text{ kHz}$  (从而使得所模拟的光伏系统可以最大程度地接近真实系统, 当然, 选择较低的  $50 \text{ kHz}$  并不会影响计算精度, 本文采用与文献[8]一致的  $100 \text{ kHz}$  采样频率). 其中, 温度和光照强度通过 DS1006 平台实时仿真测量得到, 并将数据传输到 DS1104 平台计算输出光伏电压  $V_{\text{pv}}$ . 需要说明的是, 光伏系统、温度和光照强度均置于同一 DS1006 平台, 因此三者的采样频率统一。

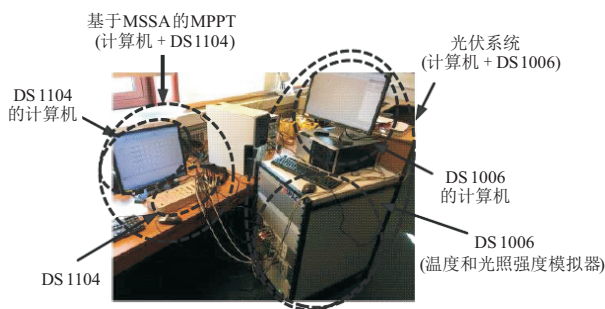


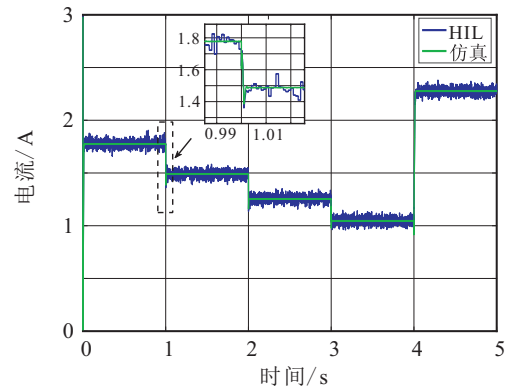
图12 HIL实验平台

#### 5.1 恒温变光照强度

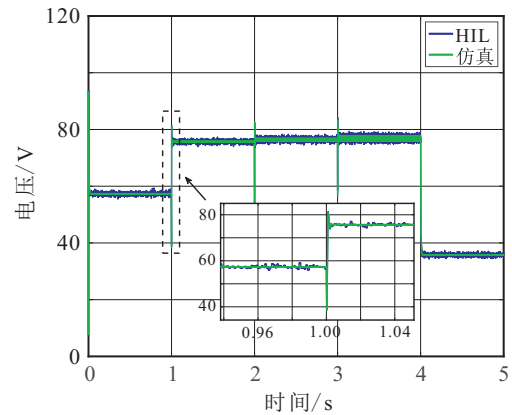
测试TRL在恒温变光照下的MPPT性能, 系统响应如图13所示. 由图13可知, HIL实验结果与仿真结果十分接近。

#### 5.2 变温变光照强度

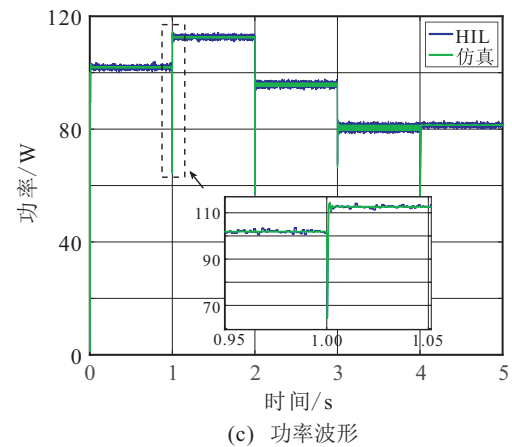
图14比较了变温变光照下仿真和HIL系统响应. 结果表明, HIL实验曲线与仿真曲线非常相似。



(a) 电流波形



(b) 电压波形



(c) 功率波形

图13 恒温变光照强度下仿真和HIL实验结果

### 6 结论

本文设计了一种全新的基于动作空间分解的TRL算法, 并成功地应用于光伏系统的MPPT. 本文的主要贡献可总结如下:

1) 通过采用动作空间分解方法, TRL可以更好地满足光伏MPPT的精确控制, 从而获得高质量的最优解, 使光伏系统在不同工况下产生更多的能量。

2) 知识迁移可为算法提供具有指导性的先验知识, 能有效避免算法在初始学习阶段的盲目搜索, 从而大幅提高收敛速率和收敛稳定性. 因此, 在不同工况下, 相比于其他算法, TRL不仅使得光伏系统输出能量最大, 而且可明显减小功率波动。

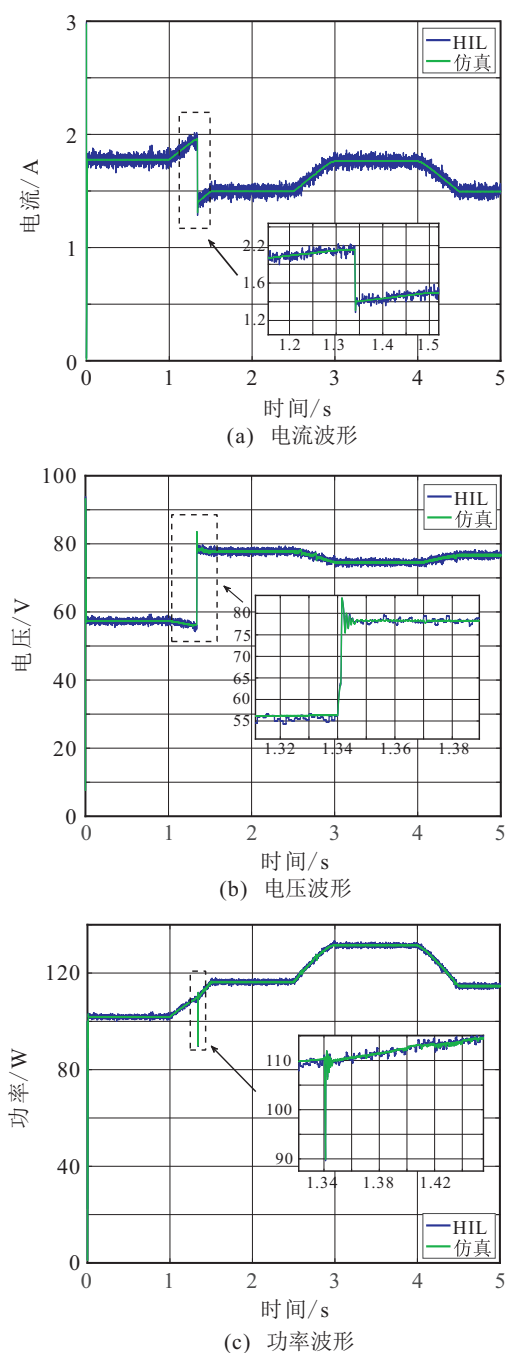


图14 变温变光照强度下仿真和HIL实验结果

#### 参考文献(References)

- [1] 周元贵, 陈启卷, 何昌炎, 等. 局部阴影下光伏阵列建模及多峰值MPPT控制[J]. 太阳能学报, 2016, 37(10): 2484-2490.  
(Zhou Y G, Chen Q J, He C Y, et al. Model of PV array under partial shading and MPPT control of multi-peak characteristics[J]. Acta Energeticae Solaris Sinica, 2016, 37(10): 2484-2490.)
- [2] Sera D, Mathe L, Kerekes T, et al. On the perturb-and-observe and incremental conductance MPPT methods for PV systems[J]. IEEE Journal of Photovoltaics, 2013, 3(3): 1070-1078.
- [3] Messai A, Mellit A, Guessoum A, et al. Maximum power

point tracking using a GA optimized fuzzy logic controller and its FPGA implementation[J]. Solar Energy, 2011, 85(2): 265-277.

- [4] Mirhassani S M, Golroodbari S Z M, Golroodbari S M M, et al. An improved particle swarm optimization based maximum power point tracking strategy with variable sampling time[J]. International Journal of Electrical Power & Energy Systems, 2015, 64: 761-770.
- [5] Pilakkat D, Kanthalakshmi S. An improved P&O algorithm integrated with artificial bee colony for photovoltaic systems under partial shading conditions[J]. Solar Energy, 2019, 178: 37-47.
- [6] Ahmed J, Salam Z. A maximum power point tracking (MPPT) for PV system using cuckoo search with partial shading capability[J]. Applied Energy, 2014, 119: 118-130.
- [7] Rezk H, Fathy A. Simulation of global MPPT based on teaching — Learning-based optimization technique for partially shaded PV system[J]. Electrical Engineering, 2017, 99(3): 847-859.
- [8] Yang B, Yu T, Zhang X S, et al. Dynamic leader based collective intelligence for maximum power point tracking of PV systems affected by partial shading condition[J]. Energy Conversion and Management, 2019, 179: 286-303.
- [9] 陈希亮, 曹雷, 李晨溪, 等. 基于重抽样优选缓存经验回放机制的深度强化学习方法[J]. 控制与决策, 2018, 33(4): 600-606.  
(Chen X L, Cao L, Li C X, et al. Deep reinforcement learning via good choice resampling experience replay memory[J]. Control and Decision, 2018, 33(4): 600-606.)
- [10] Pan J, Wang X S, Cheng Y H, et al. Multisource transfer double DQN based on actor learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(6): 2227-2238.
- [11] Taylor M E, Stone P. Transfer learning for reinforcement learning domains: A survey[J]. Journal of Machine Learning Research, 2009, 10: 1633-1685.
- [12] Pan J, Wang X S, Cheng Y H, et al. Multi-source transfer ELM-based Q learning[J]. Neurocomputing, 2014, 137: 57-64.
- [13] Taylor M E, Kuhlmann G, Stone P. Autonomous transfer for reinforcement learning[C]. Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems. Estoril: IEEE, 2008: 283-290.
- [14] 徐茂鑫, 张孝顺, 余涛. 迁移蜂群优化算法及其在无功优化中的应用[J]. 自动化学报, 2017, 43(1): 83-93.

- (Xu M X, Zhang X S, Yu J. Transfer bees optimizer and its application on reactive power optimization[J]. *Acta Automatica Sinica*, 2017, 43(1): 83-93.)
- [15] 韩传家, 张孝顺, 余涛, 等. 风险调度中引入知识迁移的细菌觅食强化学习优化算法[J]. *电力系统自动化*, 2017, 41(8): 69-77.  
(Han C J, Zhang X S, Yu T, et al. Optimization algorithm of reinforcement learning based knowledge transfer bacteria foraging for risk dispatch[J]. *Automation of Electric Power Systems*, 2017, 41(8): 69-77.)
- [16] 张孝顺, 李清, 余涛, 等. 基于协同一致性迁移Q学习算法的虚拟发电部落AGC功率动态分配[J]. *中国电机工程学报*, 2017, 37(5): 1455-1467.  
(Zhang X S, Li Q, Yu T, et al. Collaborative consensus transfer Q-learning based dynamic generation dispatch of automatic generation control with virtual generation tribe[J]. *Proceedings of the CSEE*, 2017, 37(5): 1455-1467.)
- [17] Chen K, Tian S L, Cheng Y H, et al. An improved MPPT controller for photovoltaic system under partial shading condition[J]. *IEEE Transactions on Sustainable Energy*, 2014, 5(3): 978-985.
- [18] Chen C, Takahashi T, Nakagawa S, et al. Reinforcement learning in depression: A review of computational research[J]. *Neuroscience & Biobehavioral Reviews*, 2015, 55: 247-267.
- [19] Guo H X, Liu Y Q, Wu J, et al. A reinforcement learning approach to STATCOM controller[C]. *IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies*. Hong Kong: IEEE, 2004: 638-642.
- [20] Mirjalili S. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm[J]. *Knowledge-Based Systems*, 2015, 89: 228-249.
- [21] Zhang X S, Yu T, Yang B, et al. Accelerating bio-inspired optimizer with transfer reinforcement learning for reactive power optimization[J]. *Knowledge-Based Systems*, 2017, 116: 26-38.
- [22] Bianchi R A C, Celiberto L A, Santos P E, et al. Transferring knowledge as heuristics in reinforcement learning: A case-based approach[J]. *Artificial Intelligence*, 2015, 226: 102-121.

### 作者简介

杨博(1988—), 男, 副教授, 博士, 从事新能源发电/储能系统优化与控制、人工智能在智能电网中的应用等研究, E-mail: yangbo\_ac@outlook.com;

THIDAR Swe(1990—), 女, 硕士生, 从事光伏发电系统优化与控制的研究, E-mail: 2918301836@qq.com;

钟林恩(1996—), 男, 硕士生, 从事分布式发电系统优化与控制的研究, E-mail: zhonglinen@hotmail.com;

束洪春(1961—), 男, 教授, 博士生导师, 从事电力系统新型继电保护与故障测距、故障录波、数字信号处理及DSP应用等研究, E-mail: kmshc@sina.com.cn;

张孝顺(1990—), 男, 副教授, 博士, 从事电力系统优化控制与启发式算法等研究, E-mail: xszhang1990@sina.cn;

余涛(1974—), 男, 教授, 博士生导师, 从事复杂电力系统的非线性控制理论和仿真、人工智能在电力系统中的应用等研究, E-mail: taoyu1@scut.edu.cn.

(责任编辑: 李君玲)