

# 基于强化学习的电磁悬浮型磁浮列车 悬浮控制

胡轲琰<sup>1,2</sup>, 徐俊起<sup>1</sup>, 刘志刚<sup>3</sup>, 林国斌<sup>1</sup>

(1. 同济大学 国家磁浮交通工程技术研究中心, 上海 201804; 2. 同济大学 交通运输工程学院, 上海 201804;

3. 同济大学 铁道与城市轨道交通研究院, 上海 201804)

**摘要:** 为了保证磁浮列车的安全、可靠运行,研究了悬浮系统在参数摄动条件下的悬浮控制问题。首先,对电磁悬浮(EMS)型磁浮列车的基本悬浮单元建模,给出了电流控制模型;然后,建立了悬浮系统的强化学习环境以及软演员-评论家(SAC)智能体,并设计了加速训练的奖励函数与“吸死”处理方案;最后,提出了基于强化学习的悬浮控制方法。与传统比例-积分-微分(PID)控制方法的对比结果表明,本方法具有更快的动态响应,在损失 50% 线圈匝数或磁极面积变化时具有更好的跟踪精度。

**关键词:** 电磁悬浮(EMS)型磁浮列车;悬浮控制;强化学习控制;软演员-评论家(SAC)智能体;奖励函数设计

中图分类号: U27

文献标志码: A

suspension-contact strategy were designed for accelerating the training process. Finally, a corresponding reinforcement learning method was proposed. Compared with the conventional proportional-integral-derivative (PID) control method, the proposed method has faster dynamic response and better tracking accuracy under the condition of 50% loss of coil turns and the pole area change of the magnet.

**Key words:** electromagnetic suspension (EMS) maglev trains; suspension control; reinforcement learning control; soft actor-critic (SAC) agent; reward function design

## Reinforcement Learning-based Suspension Control for Electromagnetic Suspension Maglev Trains

HU Keting<sup>1,2</sup>, XU Junqi<sup>1</sup>, LIU Zhigang<sup>3</sup>, LIN Guobin<sup>1</sup>

(1. National Maglev Transportation Engineering R&D Center, Tongji University, Shanghai 201804, China; 2. College of Transportation Engineering, Tongji University, Shanghai 201804, China; 3. Institute of Rail Transit, Tongji University, Shanghai 201804, China)

**Abstract:** In order to ensure the safe and reliable operation of maglev trains, the suspension control of suspension systems under the condition of parameter perturbation was studied. Firstly, the basic suspension unit of a electromagnetic suspension (EMS) maglev train was modeled and the current control model was given. Then, the reinforcement learning environment and the soft actor-critic (SAC) agent were established for the suspension system, and a reward function and an anti-

高速磁浮由电磁力实现非接触的支承、导向、牵引以及部分制动,并且通过长定子和非接触供电的形式获取动力和列车所需电能,可摆脱黏着利用和滑动受流这 2 个制约传统轮轨速度的主要因素,实现更高速度的运行<sup>[1]</sup>。常压条件下,高速磁浮的经济可用时速达 600 km·h<sup>-1</sup>,填补了高速铁路与航空运输之间的速度空白,为超大城市间的快速出行提供条件<sup>[2]</sup>。600 km·h<sup>-1</sup> 高速磁浮系统样机的成功联调标志着我国电磁悬浮(EMS)型高速磁浮交通研究已达到新的阶段<sup>[3]</sup>。

悬浮系统作为 EMS 型高速磁浮列车实现无接触运行的关键一环,其性能直接决定了列车的乘坐舒适度以及运行安全性。奇然等<sup>[4]</sup>提出了一种渐起渐落的策略,提高了磁浮列车起浮/降落过程的平稳性。Munderh 等<sup>[5]</sup>基于最优控制,提出了一种线性二次型调节(LQR)悬浮控制器,相比于比例-积分-微

收稿日期: 2022-12-30

基金项目: 中国博士后科学基金(2022M712408);国家自然科学基金(52232013, 52072269)

第一作者: 胡轲琰(1991—),男,工学博士,主要研究方向为磁浮列车悬浮控制。E-mail:huketing121@163.com

通信作者: 徐俊起(1977—),男,高级工程师,工学博士,主要研究方向为磁浮列车悬浮控制技术及车-轨耦合动力学。

E-mail:xujunqi@tongji.edu.cn



论文  
拓展  
介绍

分(PID)控制器,该控制器提高了系统稳定性,并且具有更快的响应速度。然而,上述方法的本质仍然是线性控制,存在着对扰动敏感的问题。为此,Ni等<sup>[6]</sup>引入了反馈线性化,对轨道齿槽效应和负载扰动进行了补偿,取得了较好的悬浮效果。孙友刚等<sup>[7]</sup>提出了基于滑模控制的悬浮控制方法,提高了系统的鲁棒性。在此基础上,靖永志等<sup>[8]</sup>根据当前状态,利用自适应控制对滑模趋近率进行在线调节,提高了系统的动态性能。佟来生等<sup>[9]</sup>将哈密顿-雅可比不等式(HJI)理论引入滑模控制,实现了在轨道不平顺激励下的悬浮间隙稳定。

针对外界干扰造成悬浮间隙波动的问题,余建国等<sup>[10]</sup>提出了基于线性自抗扰控制(LADRC)的悬浮控制器,并和PID控制和滑模控制进行了对比,结果表明LADRC具有更好的抗干扰性能。陈萍等<sup>[11]</sup>融合了滑模控制和自抗扰控制的优点,提出了滑模自抗扰悬浮控制算法,在取得较好悬浮控制效果的同时,实现了部分传感器失效下的稳定悬浮。Hu等<sup>[12]</sup>在模型预测控制的基础上考虑了状态约束,在获得较好抗干扰性能的同时优化了系统能耗。Klauco等<sup>[13]</sup>针对模型预测控制计算量大的问题,将离线分段仿射函数和二进制搜索树相结合,实现了亚毫秒级的悬浮系统模型预测控制。Zhang等<sup>[14]</sup>针对不同电感参数引起的悬浮间隙波动问题,提出了多平衡点反馈线性化方法,从而提高了系统的动态性能。Sun等<sup>[15]</sup>提出了一种幅值饱和的自适应控制方法,可准确估计悬浮点的质量,并实现位置的准确跟踪。

随着处理器的发展,越来越多的智能算法也被应用于悬浮控制。陈琛等<sup>[16]</sup>在滑模控制的基础上,引入了加速度反馈,并通过径向基函数神经网络对振动进行了抑制。王强<sup>[17]</sup>利用神经网络的逼近功能,对外界扰动进行抑制,实现了非线性负载力和不平顺扰动下的平稳悬浮。欧阳清华等<sup>[18]</sup>提出了自适应径向基函数悬浮控制方法,并验证了该方法性能优于滑模控制。Sun等<sup>[19]</sup>利用神经网络对时变的质量进行估计,并对输入延时进行补偿。Sun等<sup>[20]</sup>结合了滑模控制、自适应模糊估计器以及神经-模糊切换率,提出了一种可对参数摄动进行补偿的悬浮控制方法。

上述研究更多地关注外界扰动下控制系统的鲁棒性。在悬浮列车运行过程中,线圈匝数可能因匝间短路而变化,电磁铁磁极面积也可能因车体过弯或晃动而变化。针对上述悬浮系统参数的变化,目

前相关研究相对较少。为此,提出了一种基于强化学习的悬浮控制算法,用于减小系统参数摄动对系统性能的影响。首先,给出了悬浮系统的状态空间、动作空间以及模拟环境;其次,分析了悬浮系统环境中奖励函数对于智能体训练的影响,并提出了可实现快速训练的奖励函数;然后,基于软演员-评论家(SAC)智能体,提出了抗参数摄动的悬浮系统控制算法;最后,通过与传统方法的对比,对所提算法的有效性进行了验证。

## 1 悬浮系统建模

EMS型高速磁浮列车利用悬浮电磁铁与轨道之间的电磁吸力平衡车体重力和扰动,实现磁浮车体的稳定悬浮,悬浮系统结构如图1所示。一节列车由若干悬浮架组成,每个悬浮架由4个基本悬浮单元组成。每个悬浮单元包括一套完整的悬浮系统,由悬浮电磁铁、悬浮控制器以及悬浮传感器等组成,具体结构如图1左图所示。间隙传感器和电流传感器将间隙和电流信息反馈至控制器,通过计算得到参考电流 $i_{m,ref}$ ,由斩波器调节电磁铁电流 $i_m$ ,输出合适的电磁力,从而保持车体的悬浮。由此可见,对悬浮单元进行研究,具有一定的代表性。

假设电磁中铁磁材料的磁导率无穷,磁势均匀地降落在气隙上,并且忽略绕组的漏磁通,可以得到电磁铁的电感为

$$L(c, i) = \frac{N}{i(t)} \varphi_T = \frac{N}{i(t)} \frac{Ni(t)}{R_T} = \frac{\mu_0 N^2 A}{2c(t)} \quad (1)$$

式中: $N$ 为电磁铁绕组匝数; $\varphi_T$ 为主极磁通; $R_T$ 为磁路磁阻; $A$ 为铁芯面积; $i(t)$ 为控制电流; $\mu_0$ 为真空磁导率; $c(t)$ 为悬浮间隙。

由式(1)可知,电磁铁电感与悬浮间隙有关,是一个时变量。当悬浮间隙波动时,电感随之波动。电磁铁吸力计算式为

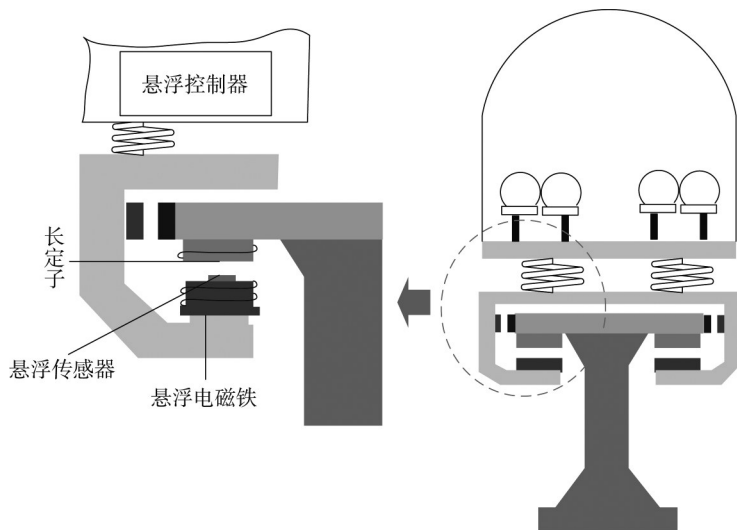
$$F_e(i, c) = \frac{B^2 A}{\mu_0} = \frac{\mu_0 N^2 A}{4} \left( \frac{i(t)}{c(t)} \right)^2 \quad (2)$$

$$B \approx \frac{\varphi_T}{A} = \frac{\mu_0 Ni(t)}{2c(t)} \quad (3)$$

可以发现,电磁铁吸力和悬浮间隙的平方成反比,即气隙越小吸力越大,这是悬浮系统不稳定的原因所在。在垂直方向上,电磁铁的运动方程为

$$m \frac{d^2 c(t)}{dt^2} = mg + f_d(t) - F_e(i, c) \quad (4)$$

式中: $m$ 为电磁铁等效质量; $f_d(t)$ 为外界干扰。

图1 EMS型高速磁悬浮系统示意图<sup>[6]</sup>Fig.1 Schematic diagram of EMS high-speed magnetic suspension system<sup>[6]</sup>

在平衡点处,有

$$mg = F_e(i_0, c_0) \quad (5)$$

式中:  $i_0$  为平衡点电流;  $c_0$  为平衡点悬浮间隙。由此,可以得到电流控制的基本悬浮单元模型,如下所示:

$$\begin{cases} F_e(i, c) = \frac{B^2 A}{\mu_0} = \frac{\mu_0 N^2 A}{4} \left( \frac{i(t)}{c(t)} \right)^2 \\ m \frac{d^2 c(t)}{dt^2} = mg + f_d(t) - F_e(i, c) \\ mg = F_e(i_0, c_0) \end{cases} \quad (6)$$

由于式(6)中存在非线性项,难以用工程中常用的线性控制方法进行控制器设计,因此在平衡点( $i_0, c_0$ )线性化展开,略去高阶项后得到

$$\begin{cases} F_e(i, c) = F_e(i_0, c_0) + K_I \Delta i(t) - K_C \Delta c(t) \\ K_C = \left( \frac{\partial F_e}{\partial c} \right) \bigg|_{(i_0, c_0)} = \frac{\mu_0 N^2 A i_0^2}{2c_0^3} \\ K_I = \left( \frac{\partial F_e}{\partial i} \right) \bigg|_{(i_0, c_0)} = \frac{\mu_0 N^2 A i_0}{2c_0^2} \end{cases} \quad (7)$$

取悬浮间隙以及电磁铁速度(向下为正)为状态量  $X = (\Delta c, \Delta \dot{c})^T$ , 电流  $\Delta i$  为系统输入, 间隙  $\Delta c$  为输出, 可以得到系统状态方程为

$$\begin{cases} \dot{X} = \begin{bmatrix} 0 & 1 \\ K_C/m & 0 \end{bmatrix} X + \begin{bmatrix} 0 \\ -K_I/m \end{bmatrix} \Delta i \\ Y = [1 \quad 0] X \end{cases} \quad (8)$$

由式(8)可知, 系统的开环极点为  $\pm \sqrt{K_C/m}$ , 存

在复平面右侧的极点, 说明该系统不稳定, 需要引入反馈控制。

## 2 基于SAC算法的悬浮控制算法设计

通过智能体与环境之间的交互学习, 强化学习方法可获取当前状态以及相应的奖励, 然后调整输出以获得最大化的奖励回报, 从而实现目标。通常, 强化学习方法为了获取回报最大化的策略, 容易陷入局部最优; 而SAC算法在奖励中加入了策略的熵, 使算法尽可能多地探索可行解, 从而获得更优解<sup>[21]</sup>。因此, 采用SAC算法对EMS型高速磁浮的悬浮系统进行控制, 提高控制系统对于参数摄动的鲁棒性。

### 2.1 马尔可夫决策过程

一个马尔可夫决策过程包含状态空间  $S$ 、动作空间  $A$ 、状态转移概率  $p(s_t | s_{t-1}, a_{t-1})$ 、奖励函数  $r(s, a) : S \times A \rightarrow R$  以及折扣因子  $\gamma$ ,  $s_t, a_t$  为当前的状态和动作。强化学习的智能体通过策略  $\pi: S \rightarrow A$  和环境进行交互, 通过优化策略使累积奖励(即回报)最大, 计算式如下所示:

$$\max_{\pi \in P} J(\pi) = \max_{\pi \in P} E \left( \sum_{t=1}^T \gamma^{t-1} r_t | \pi \right) \quad (9)$$

式中:  $P$  为策略空间;  $T$  为总步数;  $t$  为当前步数;  $r_t$  为当前奖励;  $E$  为期望。

### 2.2 SAC算法

如前所述, 通常强化学习方法以获取最大化回报为目的对策略进行学习, 即:



$$\pi^* = \arg \max_{\pi} E_{(s_t, a_t) \sim \rho_{\pi}} \left( \sum_t R(s_t, a_t) \right) \quad (10)$$

式中:  $\pi^*$  为最优策略;  $\rho_{\pi}$  为策略  $\pi$  轨迹分布的边际。然而,该算法容易集中于某一个动作,从而陷入局部最优。为了改善这一情况,可在学习过程中加入策略熵<sup>[22]</sup>,从而增加输出动作的分散程度,增强智能体的探索能力,如下所示:

$$\begin{cases} \pi^* = \arg \max_{\pi} \sum_t E_{(s_t, a_t) \sim \rho_{\pi}} \left( \underbrace{r(s_t, a_t)}_{\text{奖励函数}} + \alpha \underbrace{H(\pi(\cdot|s_t))}_{\text{策略熵}} \right) \\ H(\pi(\cdot|s_t)) = E_{(s_t, a_t) \sim \rho_{\pi}} (-\log \pi(a_t|s_t)) \end{cases} \quad (11)$$

式中:  $\alpha$  为温度系数,用于调节探索程度。可以看出,在相同奖励条件下,智能体倾向于选择熵大的策略,从而拓宽探索的范围。在最大熵条件下,贝尔曼方程  $Q(s_t, a_t)$  修正为

$$J_Q(\theta) = E_{(s_t, a_t) \sim D} \left( \frac{1}{2} (Q_{\theta}(s_t, a_t) - (r(s_t, a_t) + \gamma E_{s_{t+1}}(V_{\bar{\theta}}(s_{t+1})))^2 \right) \quad (14)$$

式中:  $D$  为经验回放池;  $\theta$  为  $Q$  值网络的参数;  $\bar{\theta}$  为目标  $Q$  值网络的参数。  $V$  值可通过式(12)利用  $Q$  值进

$$\hat{\nabla}_{\bar{\theta}} J_Q(\theta) = \nabla_{\bar{\theta}} Q_{\theta}(s_t, a_t) (Q_{\theta}(s_t, a_t) - (r(s_t, a_t) + \gamma (Q_{\bar{\theta}}(s_{t+1}, a_{t+1}) - \alpha \log(\pi_{\phi}(a_{t+1}|s_{t+1})))) \quad (15)$$

式(14)和式(15)中的  $\bar{\theta}$  由下式进行更新:

$$\bar{\theta} \leftarrow \tau \theta + (1 - \tau) \bar{\theta} \quad (16)$$

式中:  $\tau$  为 0 至 1 之间的数。

$$J_{\pi}(\phi) = E_{s_t \sim D, \epsilon_t \sim N} (\alpha \log \pi_{\phi}(f_{\phi}(\epsilon_t; s_t)|s_t) - Q_{\theta}(s_t, f_{\phi}(\epsilon_t; s_t))) \quad (18)$$

式中:  $N$  为高斯分布;  $\phi$  为策略网络的参数;  $\epsilon_t$  为高斯噪声;  $f_{\phi}$  输出均值和方差。式(17)可通过下式的近似梯度进行迭代求解:

$$\hat{\nabla}_{\phi} J_{\pi}(\phi) = \nabla_{\phi} \alpha \log(\pi_{\phi}(a_t|s_t)) + (\nabla_{a_t} \alpha \log(\pi_{\phi}(a_t|s_t)) - \nabla_{a_t} Q(s_t, a_t)) \nabla_{\phi} f_{\phi}(\epsilon_t; s_t) \quad (19)$$

为了实现温度系数的自适应调整,在新的区域扩大探索范围,在旧的区域缩小探索范围,Haarnoja 等<sup>[21]</sup>将其归纳为期望熵约束下最大化期望回报,如下所示:

$$\begin{aligned} & \max_{\pi_{0:T}} E_{\rho_{\pi}} \left( \sum_{t=0}^T r(s_t, a_t) \right) \\ & \text{s. t. } E_{(s_t, a_t) \sim \rho_{\pi}} (-\log(\pi_t(a_t|s_t))) \geq H, \forall t \end{aligned} \quad (20)$$

## 2.3 基于SAC的悬浮控制算法

### 2.3.1 智能体设定

对于悬浮系统来说,智能体的状态空间可设为各传感器的值或可从传感器推导得到的值,包括悬浮间隙、电磁铁运动速度,即  $S = \{c, \dot{c}\}$ 。根据悬浮系

$$\begin{cases} Q(s_t, a_t) \triangleq r(s_t, a_t) + \gamma E_{s_{t+1}}(V(s_{t+1})) \\ V(s_t) = E_{a_t \sim \pi}(Q(s_t, a_t) - \alpha \log \pi(a_t|s_t)) \end{cases} \quad (12)$$

式中:  $V(s_t)$  为状态价值函数。

策略的迭代更新包括策略评估以及策略更新,前者可由式(12)实现,后者由下式实现:

$$\pi_{\text{new}} = \arg \min_{\pi \in \Pi} D_{\text{KL}} \left( \pi'(\cdot|s_t) \parallel \frac{\exp\left(\frac{1}{\alpha} Q_{\pi_{\text{old}}}(s_t, \cdot)\right)}{Z_{\pi_{\text{old}}}(s_t)} \right) \quad (13)$$

式中:  $\Pi$  为策略分布的集合(本研究中为高斯分布);  $Q_{\pi_{\text{old}}}(s_t, \cdot)$  为更新前的  $Q$  函数;  $Z_{\pi_{\text{old}}}(s_t)$  为配分函数,用于归一化;  $D_{\text{KL}}$  为 Kullback-Leibler 旋度,求其最小值使得策略分布向软  $Q$  函数的指数更新。

为了保证模型的性能,采用深度神经网络对软  $Q$  函数以及策略分布进行近似。软  $Q$  网络的参数  $\theta$  可通过最小化软贝尔曼误差实现,如下所示:

行参数化。式(14)可通过随机梯度求最小值,计算式如下所示:

策略网络的参数可以通过最小化式(13)实现,利用式(17)对策略重参数化可得式(18),如下所示:

$$a_t = f_{\phi}(\epsilon_t; s_t) = f_{\phi, \mu}(s_t) + \epsilon_t \odot f_{\phi, \sigma}(s_t) \quad (17)$$

统实际情况,间隙范围为  $[c_{\min}, c_{\max}]$ ,速度范围为  $[\dot{c}_{\min}, \dot{c}_{\max}]$ 。智能体的动作为控制量  $i$ ,即  $A = \{i\}$ 。由于 SAC 算法中用  $\tanh$  函数将输出动作限制在了  $[-1, 1]$ ,因此需要将 SAC 输出的动作进行线性变换,将其映射至实际电流范围  $[i_{\min}, i_{\max}]$ 。

### 2.3.2 环境模型

环境模型为智能体提供了当前动作和状态的奖励,以及该条件下的下一状态。对于悬浮系统而言,下一状态可通过对式(6)进行离散化获取,如下所示:

$$\begin{cases} \ddot{c}(t) = g + \frac{f_d}{m} - \frac{\mu_0 N^2 A i^2(t)}{4mc^2(t)} \\ \dot{c}(t+1) = \dot{c}(t) T_s \\ c(t+1) = c(t) T_s \end{cases} \quad (21)$$

式中:  $T_s$  为系统采样时间。

由于机械结构对于电磁铁运动范围的限制,因此电磁铁的运动速度存在以下限制:

$$\dot{c} = \begin{cases} 0, c \leq c_{\min} \text{ 且 } F_e < 0 \\ 0, c \geq c_{\max} \text{ 且 } F_e > 0 \\ \dot{c}, \text{ 其他} \end{cases} \quad (22)$$

Zhao等<sup>[23]</sup>参照LQR的能量函数,同时考虑状态误差与消耗能量,使策略 $\pi$ 向设定间隙靠近,并且减少能量消耗、减小间隙波动,从而给出悬浮系统的奖励函数,如下所示:

$$r(s_t, u_t) = -(c_t - c_0)^2 - 0.1\dot{c}_t^2 - 0.05u_t^2 \quad (23)$$

由于采用电压控制方案,因此存在电压项 $u_t$ ,其余2项分别为悬浮间隙偏离项以及系统振动项。

本质上来说,式(23)是一种负奖励函数,智能体容易选择提前结束本次回合来提高奖励回报。为此,文献[23]中未设置提前结束的条件,而是设置了固定的回合步数。然而,由式(6)可知,电磁铁吸力与间隙的平方成反比,间隙越小吸力越大,未经良好训练的智能体容易发生“吸死”现象。在“吸死”现象发生后,悬浮间隙不再发生变化,此后的步数对于智能体的训练效果较差,这可能导致智能体的训练收敛较慢。因此,从正向奖励的角度设计奖励函数。对于靠近控制目标的状态和动作给予更高的奖励,反之则给予较小的奖励。如此,智能体将避免回合结束来保证回报的最大化。为了实现这一目的,采用指数函数形式的奖励函数,如下所示:

$$r(s_t, i_t) = 0.9^{1000(c_t - c_0)^2} + 0.005 \times 0.9^{c_t^2} + 0.05 \times 0.9^{i_t^2} \quad (24)$$

0.9为大于0小于1的值,可以通过调整其值来调节奖励函数对于不同误差的敏感程度。图2给出了不同底数条件下奖励函数对于间隙误差的敏感性。通过实际测试,取底数为0.9。在此种条件下,可以设置发生“吸死”条件即为回合结束,从而提高智能体训练效率。

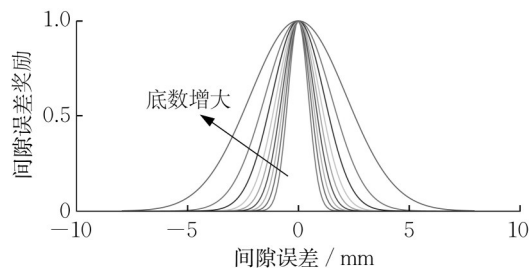


图2 指数函数形式的间隙误差奖励函数

Fig.2 Reward function of gap error with exponential function form

### 2.3.3 控制算法流程

基于SAC算法的悬浮控制框架如图3所示。Actor由包含2个全连接层的神经网络组成,神经元个数分别为400和300,激活函数为ReLU函数。得到均值 $\mu$ 和方差 $\sigma$ 之后,采样得到动作 $a_t$ ,并通过tanh函数进行幅值映射。Actor通过动作 $a_t$ 与环境进行交互,环境给出下一状态 $s_t$ 以及奖励 $r_t$ ,获取经验 $(s_t, a_t, r_t, s_{t+1})$ 并存入经验回放池 $D$ 。Critic中采用了2个Q网络以提高训练的稳定性,取2个网络输出的最小值。

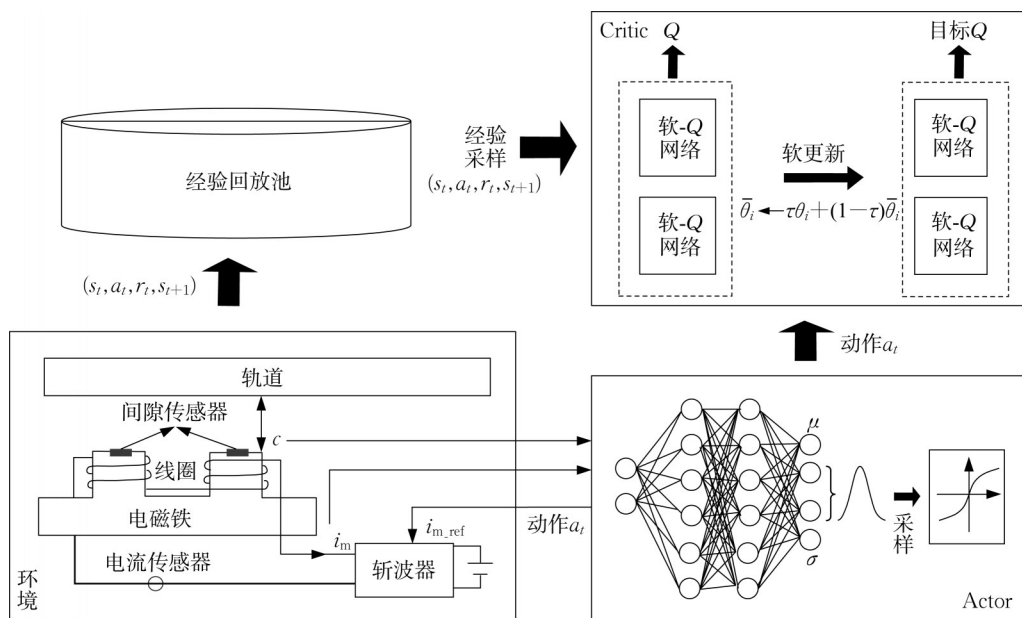


图3 基于SAC的悬浮控制框架示意图

Fig.3 Schematic diagram of SAC-based suspension control framework

控制算法包括训练和实际控制两部分,分别如图 4、5 所示。

基于SAC的悬浮控制智能体训练算法

初始化经验回放池D为空,重置环境E

随机初始化2个Q网络参数 $\theta_1$ 、 $\theta_2$

随机初始化Actor网络参数 $\phi$

利用 $\theta_1$ 、 $\theta_2$ 初始化目标网络参数 $\bar{\theta}_1$ 、 $\bar{\theta}_2$

for 每个回合 do

随机初始化环境

for 每个环境步 do

获取当前状态 $s_t$ , 抽取动作 $a_t \sim \pi(\cdot|s_t)$

在环境中执行动作 $a_t$

获取奖励 $r_t$ 、回合结束信号 $d_t$ , 以及下一状态 $s_{t+1}$

将 $(s_t, a_t, r_t, s_{t+1})$ 存入经验回放池D

若 $d_t$ 为真或大于最大步数W, 则结束当前回合

end for

end for

for 每个梯度更新步 do

从经验回放池D抽取一批 $(s_t, a_t, r_t, s_{t+1})$

更新Critic网络参数 $\theta_i \leftarrow \theta_i - \lambda_Q \nabla_{\theta_i} \lambda_Q(\theta_i)$ ,  $i \in \{1, 2\}$

更新Actor网络参数 $\phi \leftarrow \phi - \lambda_\pi \nabla_\phi J_\pi(\phi)$

更新温度系数 $\alpha \leftarrow \alpha - \lambda \nabla_\alpha J(\alpha)$

更新目标网络参数 $\bar{\theta}_i \leftarrow \tau \theta_i + (1-\tau)\bar{\theta}_i$ ,  $i \in \{1, 2\}$

end for

输出: 策略网络 $\pi$

图 4 基于SAC的悬浮控制智能体训练算法  
Fig.4 SAC-based agent training algorithm for suspension control

基于SAC的悬浮控制

输入: 策略网络 $\pi$ , 环境E

初始化经验回放池D为空, 重置环境E

for 每个环境步 do

获取当前状态 $s_t$ , 并选择动作 $a_t \sim \pi(\cdot|s_t)$

对环境施加动作

获取奖励 $r_t$ 、回合结束信号 $d_t$

若 $d_t$ 为真或大于最大步数W, 则结束控制

end for

输出: 控制结果

图 5 基于SAC的悬浮控制  
Fig.5 SAC-based suspension control

3 算法验证与分析

为了验证基于SAC的悬浮控制方法的性能,将该方法与传统的PID控制分别在损失50%线圈和过弯道时磁极面积变化下的控制性能进行对比。

3.1 智能体训练结果

设置总环境更新步数M为350 000次,单回合最大步数W为1 000,采样/控制周期为0.001 s,系统参数如表1所示。为了表明所提奖励函数对于提高

训练效率的有效性,分别在式(23)(电压项修改为电流项)和式(24)奖励函数条件下对智能体进行多次训练,结果分别如图6、7所示。可以发现:式(23)奖励函数条件下,在训练步数达到最大值 $35 \times 10^4$ 之后,算法仍然没有收敛;而在式(24)奖励函数条件下,算法在 $22 \times 10^4$ 处基本都实现收敛。相比于式(23)奖励函数,式(24)奖励函数可极大地提高智能体的训练效率,节约训练时间。

表 1 系统参数  
Tab.1 System parameters

物理参数	参数值
悬浮块质量/kg	750
线圈匝数	270
目标间隙/m	0.008
电磁铁面积/m <sup>2</sup>	0.115
真空磁导率/(H·m <sup>-1</sup> )	$4\pi \times 10^{-7}$
线圈电阻/ $\Omega$	3.84
等效负载/kN	24.5

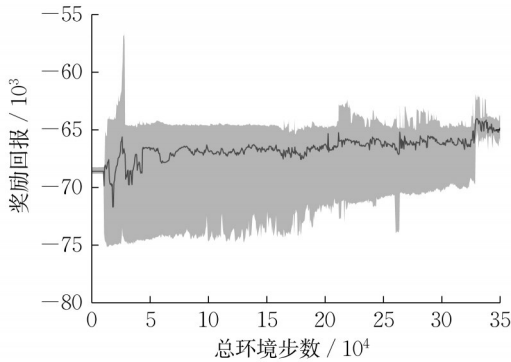


图 6 式(23)奖励函数下的每回合累积回报  
Fig.6 Return of each episode with reward function (23)

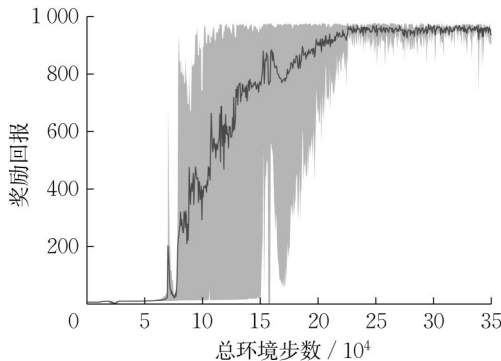


图 7 式(24)奖励函数下的每回合累积回报  
Fig.7 Return of each episode with reward function (24)

3.2 与传统算法的对比较证

为了验证基于SAC悬浮控制方法的抗参数摄

动性能,将该方法与传统PID控制在各类情况下的控制性能进行对比。依据模型(7)以及文献[16]的参数整定方法,所用的PID参数为 $k_p=20\ 000$ , $k_i=1\ 200$ , $k_d=1\ 000$ 。为了更加真实地反映运行条件,仿真采用式(6)的非线性模型,而非式(7)的线性化模型。在以下所有仿真中,都对电磁铁施加 $[-10\ 000,10\ 000]$  N的随机干扰力。

### 3.2.1 静态起浮

图8为静态起浮时PID控制和SAC控制下的间隙误差和控制电流。从图8a可以看出,SAC控制和PID控制均有较好的控制效果,可以达到较为平稳的悬浮状态。相较而言,SAC控制比PID控制具有更快的动态响应,SAC控制在0.05 s时即可到达平稳悬浮状态,而PID控制则需花费0.24 s左右。由于随机干扰的存在,电磁铁电流需要进行调整,使悬浮间隙保持稳定。可以发现,SAC控制的电流相比于PID控制的电流波动更小,平稳性更好。

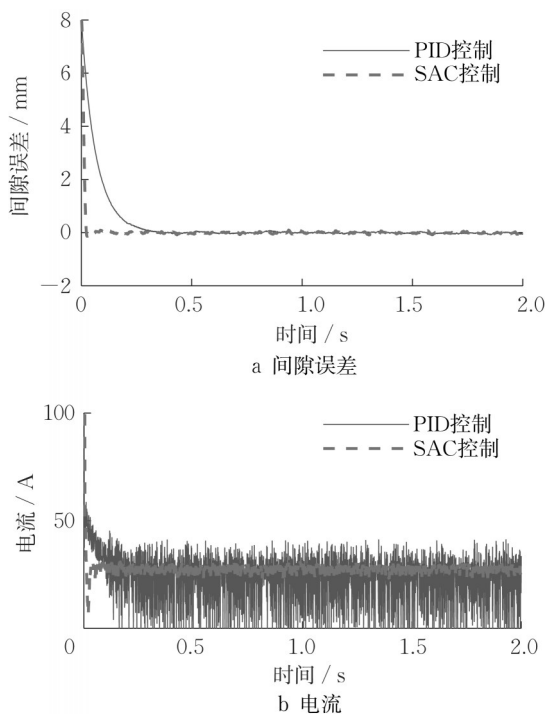


图8 静态起浮时SAC与PID控制性能对比

Fig.8 Comparison of performance between SAC and PID in term of static suspension

### 3.2.2 电磁铁线圈匝数变化

图9为悬浮条件下发生电磁铁线圈短路故障,有效线圈匝数从270变为135时PID控制和SAC控制的间隙误差和控制电流。当故障发生后,SAC控制下的间隙误差为0.96 mm,PID控制下的间隙误差为2.45 mm,PID控制的间隙误差相较于SAC控

制大了1.49 mm。悬浮间隙存在限制且相对较小,大的间隙误差增加了电磁铁触轨的风险。从图9b可以看出,当有效线圈匝数减小之后,因为PID控制的间隙误差较大,其悬浮所需的电流也相应增大,从而增加了悬浮系统的能耗。

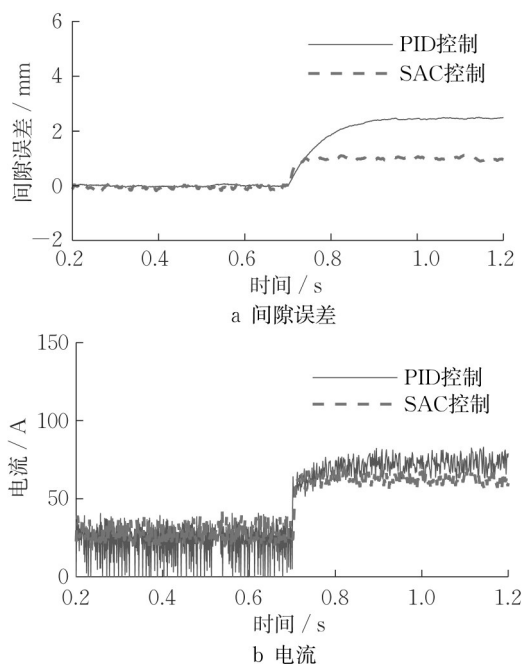


图9 线圈匝数改变时SAC与PID控制性能对比

Fig.9 Comparison of performance between SAC and PID in term of coil turn change

### 3.2.3 电磁铁磁极面积变化

磁浮列车运行过程中,在弯道时电磁铁磁极面积将发生变化。图10为磁极面积从0.1150 m<sup>2</sup>在1 s内逐步减小至0.0575 m<sup>2</sup>时PID控制和SAC控制的间隙误差和控制电流。

由图10a可知:磁极面积的变化对于SAC控制下的悬浮间隙影响不是很大,仅产生了0.3 mm左右的间隙误差;而对于PID控制下的悬浮间隙而言,间隙误差达到了1 mm,相比SAC控制增加了0.7 mm。SAC控制的间隙误差只有PID控制的30%。图8b中两者的电流相差相对较小,这是因为两者的绝对误差相对较小。

综合上述,SAC控制比PID控制具有更好的平稳性并且具备更快的响应速度。对于电磁铁线圈匝数减半、磁极面积减半等情况,相当于引入了较大的参数变动甚至是故障状态,使得被控系统与原系统产生了较大的差异,控制器难以实现对参考间隙的准确跟踪。此种条件下,SAC算法相比PID算法具有更好的性能,可在一定程度上减小间隙误差。



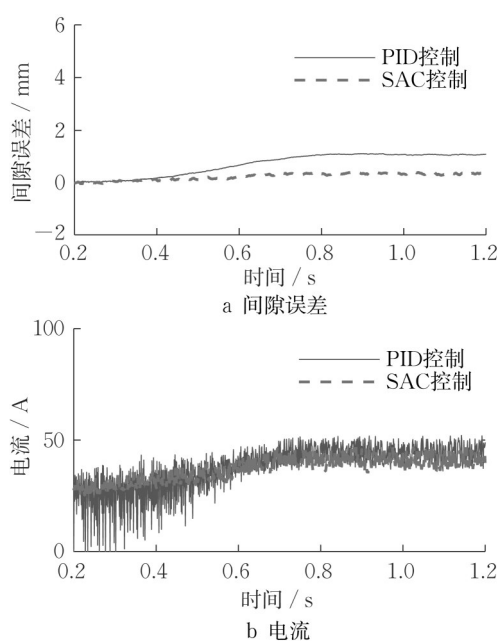


图10 磁极面积变化时SAC与PID控制性能对比

Fig.10 Comparison of performance between SAC and PID in term of pole area change

## 4 结语

针对磁浮列车动态条件或者故障条件下参数变化引起的传统悬浮控制性能下降问题,提出了一种基于强化学习的悬浮控制方法。建立了悬浮系统的强化学习环境,利用SAC算法的全局探索能力减小了悬浮系统在不同系统参数下的间隙跟踪误差。此外,针对悬浮系统存在的“吸死”现象,提出了奖励函数,提高了强化学习智能体的训练效率。与传统的PID控制算法相比,所提的SAC控制算法具有更优越的性能。SAC算法的起浮时间仅为PID算法的20.83%;在50%电磁铁线圈匝数损失时,SAC算法的间隙误差仅为PID算法的39.18%;在磁极面积缓慢变化减小时,SAC算法的间隙误差为PID算法的30%。由于目前的悬浮控制器算力性能无法支持所提的SAC悬浮控制算法的实时在线实施,下一阶段将对网络模型的小型化进行研究,从而实现基于SAC的在线悬浮控制。

### 作者贡献声明:

胡轲琰:设计、编写算法及撰写论文。

徐俊起:模型构建及数据支持。

刘志刚:提供研究思路。

林国斌:技术指导及论文审定。

### 参考文献:

- [1] 熊嘉阳,邓自刚.高速磁悬浮轨道交通研究进展[J].交通运输工程学报,2021,21(1):177.  
XIONG Jiayang, DENG Zigang. Research progress of high-speed maglev rail transit [J]. Journal of Transit and Transportation Engineering, 2021, 21(1): 177.
- [2] 余浩伟,寇俊瑜,李艳.600 km/h 高速磁浮在国内的适应性及工程化发展[J].铁道工程学报,2020,37(12):16.  
YU Haowei, KOU Junyu, LI Yan. Adaptability and engineering development of 600 km/h high-speed maglev in China [J]. Journal of Railway Engineering Society, 2020, 37(12): 16.
- [3] 梁建英.中国高速磁浮交通系统发展现状与展望[J].科学,2022,74(5):31.  
LIANG Jianying. Current status and outlook of high-speed magnetic levitation transportation system development in China [J]. Science, 2022, 74(5): 31.
- [4] 齐然,罗京.中低速磁浮列车悬浮控制系统研究[J].电力机车与城轨车辆,2016,39(3):29.  
QI Ran, LUO Jing. Suspension control system research for middle-low maglev trains [J]. Electric Locomotives & Mass Transit Vehicles, 2016, 39(3): 29.
- [5] MUNDERH Y, HADER A. Modeling and control for a magnetic levitation system based on SIMLAB platform in real time [J]. Results in Physics, 2018, 8: 153.
- [6] NI Fei, MU Siyuan, KANG Jinsong, *et al.* Robust controller design for maglev suspension systems based on improved suspension force model [J]. IEEE Transactions on Transportation Electrification, 2021, 7(3): 1765.
- [7] 孙友刚,李万莉,林国斌,等.低速磁浮列车悬浮系统动力学建模及非线性控制[J].同济大学学报(自然科学版),2017,45(5):741.  
SUN Yougang, LI Wanli, LIN Guobin, *et al.* Dynamic modeling and nonlinear control research on magnetic suspension systems of low-speed maglev train [J]. Journal of Tongji University (Natural Science), 2017, 45(5): 741.
- [8] 靖永志,冯伟,王森,等.基于自适应非奇异终端滑模的悬浮控制策略[J].西南交通大学学报,2022,57(3):566.  
JING Yongzhi, FENG Wei, WANG Sen, *et al.* Levitation control strategy based on adaptive non-singular terminal sliding mode [J]. Journal of Southwest Jiaotong University, 2022, 57(3): 566.
- [9] 佟来生,张文跃,郑文文,等.HJI理论下滑模鲁棒控制对磁浮列车悬浮特性影响研究[J].机械科学与技术,2021,40(9):1439.  
TONG Laisheng, ZHANG Wenye, ZHENG Wenwen, *et al.* Influence of HJI theory sliding mode robust control on suspension characteristics of maglev train [J]. Mechanical Science and Technology for Aerospace Engineering, 2021, 40(9): 1439.
- [10] 余建国,木柏林.基于LADRC的悬浮控制优化研究[J/OL].



- 铁道科学与工程学报. DOI: 10.19713/j.cnki.43-1423/u.T20220069.
- YU Jianguo, MU Bailin. Research on optimization of suspension control based on LADRC [J/OL]. Journal of Railway Science and Engineering. DOI: 10.19713/j.cnki.43-1423/u.T20220069.
- [11] 陈萍, 史天成, 于明月, 等. 中低速磁浮列车滑模自抗扰悬浮控制算法[J/OL]. 铁道科学与工程学报. DOI: 10.19713/j.cnki.43-1423/u.T20220349.
- CHEN Ping, SHI Tiancheng, YU Mingyue, *et al.* Sliding mode active disturbance rejection levitation control algorithm of the medium and low speed maglev vehicles [J/OL]. Journal of Railway Science and Engineering. DOI: 10.19713/j.cnki.43-1423/u.T20220349.
- [12] HU Wenjie, ZHOU Yonghua, ZHANG Zhenlin, *et al.* Model predictive control for hybrid levitation systems of maglev trains with state constraints [J]. IEEE Transactions on Vehicular Technology, 2021, 70(10): 9972.
- [13] KLAUCO M, KALUZ M, KVASNICA M. Real-time implementation of an explicit MPC-based reference governor for control of a magnetic levitation system [J]. Control Engineering Practice, 2017, 60: 99.
- [14] ZHANG Liwei, ZHANG Yue, ZHANG Chao, *et al.* Research on the improvement of feedback linearization control in suspension system countering inductance variation [J]. Mathematical Problems in Engineering, 2019, 2019(2): 1.
- [15] SUN Ning, FANG Yongchun, CHEN He. Tracking control for magnetic-suspension systems with online unknown mass identification[J]. Control Engineering Practice, 2017, 58: 242.
- [16] 陈琛, 徐俊起, 林国斌, 等. 具有径向基网络加速度反馈的磁浮列车悬浮系统滑模控制[J]. 同济大学学报(自然科学版), 2021, 49(12): 1642.
- CHEN Chen, XU Junqi, LIN Guobin, *et al.* Sliding mode control of maglev train suspension system with neural network acceleration feedback [J]. Journal of Tongji University (Natural Science), 2021, 49(12): 1642.
- [17] 王强. 基于神经网络逼近的磁浮列车动态悬浮控制[J]. 科学技术与工程, 2021, 21(28): 12117.
- WANG Qiang. Dynamic levitation control of maglev train based on neural network approximation [J]. Science Technology and Engineering, 2021, 21(28): 12117.
- [18] 欧阳清华, 樊宽刚, 雷爽. 磁悬浮球的自适应径向基函数控制方法研究[J/OL]. 铁道科学与工程学报. DOI: 10.19713/j.cnki.43-1423/u.T20220727.
- OUYANG Qinghua, FAN Kuangang, LEI Shuang. Research on adaptive radial basis function control method of magnetic levitation ball [J/OL]. Journal of Railway Science and Engineering. DOI: 10.19713/j.cnki.43-1423/u.T20220727.
- [19] SUN Yougang, XU Junqi, CHEN Chen, *et al.* Reinforcement learning-based optimal tracking control for levitation system of maglev vehicle with input time delay[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1.
- [20] SUN Yougang, XU Junqi, QIANG Haiyan, *et al.* Adaptive neural-fuzzy robust position control scheme for maglev train systems with experimental verification[J]. IEEE Transactions on Industrial Electronics, 2019, 66(11): 8589.
- [21] HAARNOJA T, ZHOU A, HARTIKAINEN K, *et al.* Soft actor-critic algorithms and applications[J/OL]. [2022-10-30]. <https://doi.org/10.48550/arXiv.1812.05905>.
- [22] 乔骥, 王新迎, 张擎, 等. 基于柔性执行器-评判器深度强化学习的电-气综合能源系统优化调度[J]. 中国电机工程学报, 2021, 41(3): 819.
- QIAO Ji, WANG Xinying, ZHANG Qing, *et al.* Optimal dispatch of integrated electricity-gas system with soft actor-critic deep reinforcement learning [J]. Proceedings of the CSEE, 2021, 41(3): 819.
- [23] ZHAO Feiran, YOU Keyou, SONG Shiji, *et al.* Suspension regulation of medium-low-speed maglev trains via deep reinforcement learning [J]. IEEE Transactions on Artificial Intelligence, 2021, 2(4): 341.