

改进 TD3 算法在电机 PID 控制器中的应用^①



张梦杰, 陈姚节, 邓江

(武汉科技大学 计算机科学与技术学院, 武汉 430081)

通信作者: 陈姚节, E-mail: chenyaojie@wust.edu.cn

摘要: 本文针对永磁同步电机 (PMSM) 在工业领域中的多变量、非线性和强耦合特性以及传统 PID 控制在应对其变化时可能导致参数调整困难、响应延迟、鲁棒性差以及适应性问题, 提出一种结合双延迟深度确定性策略梯度 (TD3) 算法和 PID 控制的创新方案, 以优化 PID 参数调整, 实现更精确的电机速度控制。在本文的方法中, 融合双向长短期记忆网络 (BiLSTM) 到 Actor 和 Critic 网络中, 极大增强了对 PMSM 动态行为的时间序列数据处理能力, 使得系统不仅能够准确捕捉当前状态, 还能预测未来趋势, 从而实现对 PID 参数更精确和适应性强的自整定。此外, 通过整合熵正则化和好奇心驱动的探索方法进一步增强策略的多样性, 避免过早收敛到次优策略, 并鼓励模型对未知环境进行深入探索。为验证方法的有效性, 设计了一个永磁同步电机的仿真模型, 并将本文提出的方法 BiLSTM-TD3-ICE 与传统的 TD3 以及经典的 Ziegler-Nichols (Z-N) 方法进行对比。实验结果充分证明了本文提出的策略在控制性能上具有显著的优势。

关键词: 深度强化学习; TD3 算法; 永磁同步电机; PID 参数自整定; BiLSTM; 内在好奇心; 熵正则化

引用格式: 张梦杰, 陈姚节, 邓江. 改进 TD3 算法在电机 PID 控制器中的应用. 计算机系统应用, 2024, 33(5):262-270. <http://www.c-s-a.org.cn/1003-3254/9498.html>

Application of Improved TD3 Algorithm in Motor PID Controllers

ZHANG Meng-Jie, CHEN Yao-Jie, DENG Jiang

(School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081, China)

Abstract: This study analyzes the multivariate, nonlinear, and strong coupling characteristics of permanent magnet synchronous motors (PMSM) in industrial applications, as well as the difficulties in their parameter adjustment, response delay, poor robustness, and adaptability issues encountered with traditional PID control. A novel approach combining a twin delayed deep deterministic policy gradient (TD3) algorithm with PID control is proposed to optimize PID parameter adjustment for more accurate motor speed control. In this method, bidirectional long short-term memory networks (BiLSTM) are integrated into the Actor and Critic networks, significantly enhancing the processing capability for time-series data of PMSM's dynamic behavior. This enables the system to accurately capture the current state and predict future trends, achieving more precise and adaptive self-tuning of PID parameters. Moreover, the integration of entropy regularization and curiosity-driven exploration methods further enhances the diversity of the strategy, preventing premature convergence to suboptimal strategies and encouraging in-depth exploration of unknown environments. To validate the effectiveness of the proposed method, a simulation model of a PMSM is designed, and the proposed BiLSTM-TD3-ICE method is compared with the traditional TD3 and the classical Ziegler-Nichols (Z-N) method. The experimental results demonstrate the significant advantages of the proposed strategy in control performance.

Key words: deep reinforcement learning; TD3 algorithm; permanent magnet synchronous motor (PMSM); PID parameter self-tuning; BiLSTM; intrinsic curiosity; entropy regularization

① 基金项目: 装备发展部“慧眼行动”项目 (62602010214)

收稿时间: 2023-11-28; 修改时间: 2023-12-29; 采用时间: 2024-01-05; csa 在线出版时间: 2024-04-01

CNKI 网络首发时间: 2024-04-03

永磁同步电机 (permanent magnet synchronous motor, PMSM) 因其高功率密度和稳定性, 在机器人、船舶、新能源汽车以及航空航天等领域得到了广泛应用^[1-4]。尽管如此, 由于其内部模型的复杂性, 精确地控制 PMSM 的运转速度仍然是一个技术难题^[5]。

PID 控制器因其简洁和易操作的特性, 在实际应用中得到广泛使用, 但参数调整仍是个挑战, 不恰当的参数配置可能会引起系统的不稳定或震荡。众多的研究已经探索如何通过集成智能控制技术来增强 PID 的性能。Ab Ghani 等人^[6]研究通过融合模糊控制优化 PID, 提高了控制效果和抗干扰性。Quang 等人^[7]进行的研究将多层感知器神经网络用于 PID 自整定, 展现出良好的鲁棒性和适应性。在 Sanatel 等人的研究中^[8], 针对非线性系统的轨迹跟踪控制问题, 将长短期记忆网络 (LSTM) 应用于 PID 控制器的参数更新中, 利用 LSTM 对 PID 控制器的 K_P 、 K_I 和 K_D 参数进行更新, 将 LSTM 良好的估计特性与鲁棒实用的 PID 控制器结合。但模糊控制依赖专家经验, 神经网络需要大量训练数据, 并且, 这些方法依赖系统模型和规则调整, 遇到突发干扰可能难以实时适应, 增加不稳定风险。

近年来随着人工智能领域的飞速进展, 深度强化学习作为其核心子领域受到了广泛关注。这种技术主要通过让代理与环境互动, 从数据中学习并优化决策策略。PID 控制参数的调整被视为黑盒优化问题, 深度强化学习为 PID 参数整定提供了新方向^[9]。Shi 等人^[10]使用 Q-learning 的自适应 PID 控制器调整控制增益, 但其不适用于连续状态-动作空间。Liao 等人^[11]则结合 Actor-Critic 算法和增量式 PID, 用 BP 神经网络近似策略与价值函数, 取得良好性能, 但由于计算延迟问题, 无法做到实时调整。Lu 等人^[12]利用 DDPG 的 PID 策略实时补偿 PID 参数, 但 DDPG 对超参数敏感, 并易过估计值函数。Fujimoto 等人^[13]提出了 TD3 算法, 引入了双 Q 网络 (twin Q-networks) 的概念, 对两个 Q 网络的预测取最小值来抑制过估计。由于 TD3 算法的优越性, Shi 等人^[14]将 TD3 算法与 PID 控制结合, 实现 PID 参数的动态整定。为了改进 DDPG 算法的稳定性, TD3 在策略网络的输出上添加噪声以鼓励探索。然而, 由于噪声的加入, 算法有时会对状态或动作的边界进行过度探索, 尤其是在动作空间有限制的情境下, 导致出现局部最优解。

本文以永磁同步电机作为研究对象, 使用改进的

TD3 算法对电机的 PID 控制器的参数进行自整定, 本文针对电机速度控制的时序特征, 在 Actor 网络和 Critic 网络中引入 BiLSTM 以实现对状态序列的双向分析, 从而使网络能够综合考虑历史和未来信息, 显著提高控制策略的精确性和鲁棒性。此外, 研究还提出一种结合熵正则化和好奇心驱动的探索策略来加强探索能力。通过引入熵正则化保持一定的探索性, 从而避免过早地收敛到一个确定性的策略, 并利用好奇心驱动的内在奖励机制来激励代理探索环境中未知的状态空间。为了验证所提方法的有效性, 研究构建了永磁同步电机的仿真模型, 并将改进后的 TD3 算法、传统的 TD3 算法以及 Z-N 方法自整定 PID 参数进行了比较, 实验结果证明, 本研究提出的方法不仅提升了探索能力, 还能够以更稳定、更高效的工作方式寻找到最佳控制策略, 确保永磁同步电机的最优运行。

1 永磁同步电机模型分析

永磁同步电机的速度控制系统采用双闭环设计, 其中内环为电流环主要用于对电流的控制, 外环为速度环专注于对速度的控制^[15]。永磁同步电机系统框架如下图所示。在此系统中, 速度环通过 PID 控制器来调节电机速度, 输入参考速度 ω_{ref} , 反馈速度为 ω_t , 系统通过计算参考速度与反馈速度的误差, 将其输入 PID 控制器, 并将其输出用作 i_q 轴的输入信号, 同时将 i_d 轴的参考输入设定为 0。 i_q 轴和 i_d 轴都采用 PID 控制器, 通过 PMSM 模块计算出电机速度。为建立永磁同步电机的数学模型, 本文基于以下假设。

- (1) 电机定子绕组的三相在空间上对称分布。
- (2) 空间磁场呈正弦分布。
- (3) 不计磁滞和涡流损耗。
- (4) 忽略铁心饱和。
- (5) 忽略齿槽转矩及其他扰动影响。

根据上述的假设, 通常选择同步旋转坐标系 D-Q 下, $i_d = 0$ 的数学模型, 其定子电压方程表示为:

$$\begin{cases} u_d = Ri_d + L_d \frac{d}{dt} i_d - \omega_e L_e i_q \\ u_q = Ri_q + L_q \frac{d}{dt} i_q + \omega_e (L_d i_d + \psi_f) \end{cases} \quad (1)$$

其中, u_d 、 u_q 分别是定子电压的 D-Q 轴分量; i_d 、 i_q 分别是定子电流的 D-Q 轴分量; R 是定子的电阻, ψ_d 、 ψ_q 为定子磁链的 D-Q 轴分量; ω_e 是电角速度; L_d 、

L_q 分别是 D-Q 轴电感分量; ψ_f 代表永磁体的磁链。

电磁转矩方程可以写为:

$$T_e = \frac{3}{2} p_n i_q [i_d (L_d - L_q) + \psi_f] \quad (2)$$

电机机械运动方程表示为:

$$J \frac{d\omega_m}{dt} = T_e - T_L - B\omega_m \quad (3)$$

电机参数如表 1 所示。永磁同步电机控制原理图如图 1 所示, 其中 $30/\pi$ 为角速度至转速的转换操作。

表 1 永磁同步电机参数

参数名	数值
极对数 p_n	4
定子电感 L_d	8.5 mH
定子电感 L_q	8.5 mH
定子电阻 R	2.875 Ω
磁链 ψ_f	0.175 Wb
转动惯量 J	0.001 $\text{kg} \cdot \text{m}^2$

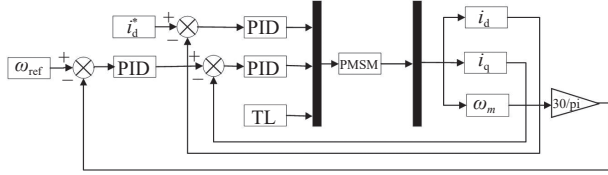


图 1 基于 PID 控制的永磁同步电机结构图

2 PID 控制与深度强化学习

2.1 PID 控制

PID 控制器之所以被称为 PID 控制器, 是因为它的输出是 3 个变量的总和: 比例项 K_P 、积分项 K_I 和导数项 K_D 。这些参数依赖于输入和输出之间的误差项 $e(t)$ 。PID 的输出 $u(t)$ 如式 (4) 描述^[16]。

$$u(t) = K_P \times e(t) + K_I \times \int_0^t e(t) dt + K_D \times \frac{de(t)}{dt} \quad (4)$$

2.2 TD3

强化学习 (reinforcement learning, RL) 是机器学习的一个分支, 核心思想在于学习如何采取最佳的行动策略, 使得智能体在与环境的交互中能够获得最大的累积奖励 $\max R(t)$ ^[17]。

$$R(t) = \sum_{i=0}^T \gamma^i r_{t+1}, \gamma \in (0, 1] \quad (5)$$

其中, γ 表示为奖励折扣因子。

为了更好地描述和解决强化学习问题, 通常将强

化学习问题表示为马尔可夫决策过程 (Markov decision process, MDP), 其中 S 为状态集、 A 为动作集、 R 为奖励函数, P 为在状态 s 下采取动作 a , 转到下一个状态 s' 的概率。

策略 (policy) 定义了在一定状态下智能体应该采取的动作的概率。具体来说, 策略通常表示为: $\pi: s \rightarrow a$, 其中 s 是状态, a 是动作。

$$\pi(a | s) = P(A_t = a | S_t = s) \quad (6)$$

为了评估策略的好坏, 引入状态价值函数 $V_\pi(s)$ 和动作价值函数 $Q_\pi(s, a)$ 。其中状态价值函数和动作价值函数使用 Bellman 方程表示如下:

$$\begin{cases} V_\pi(s) = \mathbb{E}_\pi(R_{t+1} + \gamma V_\pi(S_{t+1}) | S_t = s) \\ Q_\pi(s, a) = \mathbb{E}_\pi(R_{t+1} + \gamma Q_\pi(S_{t+1}, A_{t+1}) | S_t = s, A_t = a) \end{cases} \quad (7)$$

TD3 算法也是 Actor-Critic (AC) 框架下的一种确定性深度强化学习算法, 它结合了深度确定性策略梯度算法和双重 Q 学习, 主要是在 DDPG 算法的基础上提出了 3 个关键技术^[18]。

(1) 双重网络 (double network): 采用两个 Critic 网络, 计算目标值时取二者中的较小值, 从而抑制网络过估计问题。

(2) 目标策略平滑正则化 (target policy smoothing regularization): 计算目标值时, 在下一个状态的动作上加入扰动, 从而使得价值评估更准确。

(3) 延迟更新 (delayed update): Critic 网络更新多次后, 再更新 Actor 网络, 从而保证 Actor 网络的训练更加稳定。

为了计算期望的未来奖励, 首先使用 Actor 目标网络估算下一个状态 s' 下的动作:

$$a' = \mu'(s' | \theta^{a'}) \quad (8)$$

为了增加算法的探索性, 在动作上加上噪声:

$$a' = a' + \varepsilon (\varepsilon \sim \text{clip}(N(0, \sigma), -c, c)) \quad (9)$$

TD3 算法使用两个 Critic 目标网络来估算动作的 Q 值。为了避免过度估计, 选择两者中的最小值作为目标 Q 值。

$$y = r + \gamma \min_{i=1,2} Q'_i(s', a' | \theta_i^{Q'}) \quad (10)$$

然后利用梯度下降算法最小化评估值和目标值之间的误差 L_{C_i} , 对 Critic 网络的参数进行更新:

$$L_{c_i} = (Q_i(s, a | \theta_i^{Q'}) - y)^2 \quad (i = 1, 2) \quad (11)$$

在 Critic1 和 Critic2 网络更新 d 步之后, Actor 网络更新. 首先利用 Actor 网络计算出状态 s 下的动作.

$$a_{\text{new}} = \mu(s | \theta^\mu) \quad (12)$$

然后利用 Critic1 或者 Critic2 网络来计算状态动作对 (s, a_{new}) 的评估值, 假定使用 Critic1 网络:

$$q_{\text{new}} = Q_1(s, a_{\text{new}} | \theta^{Q_1}) \quad (13)$$

最后, 为了确保学习的稳定性, TD3 采用了“软更新”的策略, 使用一个小的权重因子 $\tau \in (0, 1)$, 将当前网络和目标网络进行加权平均, 确保了更新的平滑性和稳定性.

$$\theta^{Q'_i} = \tau \theta^{Q_i} + (1 - \tau) \theta^{Q'_i} \quad (i = 1, 2) \quad (14)$$

$$\theta^\mu = \tau \theta^\mu + (1 - \tau) \theta^\mu \quad (15)$$

3 改进 TD3-PID

3.1 整体框架

PMSM 的控制涉及对时间序列数据的处理, 其中电机的当前状态不仅取决于当前的输入, 还受到过去状态的影响. 本文在传统的 TD3 算法中将 BiLSTM 引入 Actor 网络和 Critic 网络, 能有效处理这种时间序列数据, 同时还引入了好奇心和熵正则化的联合探索策略, 本文提出的 BiLSTM-TD3-ICE 算法自整定 PID 参数的整体框架图如图 2 所示.

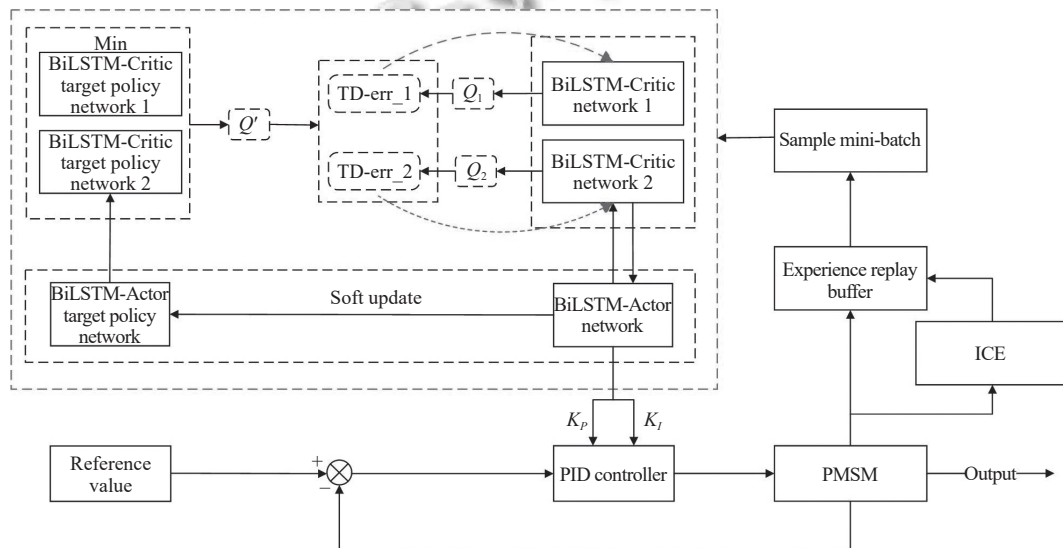


图 2 改进的 TD3-PID 整体框架图

3.2 状态、动作、奖励函数定义

本文将永磁同步电机的速度环控制问题表述为 MDP 模型, 对环境的状态空间、动作空间和奖励函数进行设置, 旨在提高电机的抗干扰能力和跟踪精度.

状态空间包括电机的实际运行速度 ω_t , 速度误差 e_t 以及 PID 控制参数的变化情况, 不仅捕捉了电机运行的即时特性, 还反映了控制器调整的动态过程. 状态 s 具体可以表示为:

$$s_t = \{\omega_t, e_t, K_P, K_I\} \quad (16)$$

动作空间定义为 PID 参数, 即比例增益 K_P 、积分增益 K_I 两个参数, 这一定义基于 PMSM 控制特性考虑, 通常只需要优化这两个参数.

$$a = \{K_P, K_I\} \quad (17)$$

奖励函数综合考虑了速度误差、电机的健康状态和温度. 这样的设计不仅可以最小化速度误差, 而且可以保持电机在一个良好的健康状态. 具体来说, 奖励计算包含以下部分.

(1) 速度误差 (e_t): 参考速度 (ω_{ref}) 和当前速度 (ω_t) 之间的差值.

$$e_t = \omega_{\text{ref}} - \omega_t \quad (18)$$

(2) 电机健康状态 (health): 该参数反映了电机的健康状态, 考虑电机健康状况超出特定阈值时的影响, 其惩罚值 (H_P) 由以下公式确定.

$$H_P = (1.0 - h) \times 0.1 \quad (19)$$

其中, h 表示电机的健康状况, 是一个介于 0-1 之间的值, 其中 0 代表最差的健康状态, 而 1 代表最佳的健康状态. 若温度超过特定阈值, h 值会相应减少, 模拟电机因高温而逐渐损坏. 这一机制确保了强化学习算法在优化控制策略时, 同时考虑电机的长期健康.

(3) 电机温度 (temperature): 表示电机在每个时间步的实时温度. 电机温度惩罚 (T_P) 公式为:

$$T_P = |T_{\text{actual}} - T_{\text{ideal}}| \quad (20)$$

综合这些误差项, 得到综合误差:

$$\mu = 0.9 \times |e(t)| + 0.1 \times |u(t)| - H_P + T_P \quad (21)$$

最终奖励使用以下方程计算:

$$\text{reward} = \exp\left(-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2\right) + 0.01 \times r \quad (22)$$

其中, σ 是一个为 10 的固定值, 用于控制奖励函数的形状. r 是一个取值范围在 [0, 1) 之间的随机数.

3.3 BiLSTM-TD3

双向长短时记忆网络 (bidirectional long short-term memory), 简称 BiLSTM, 是一种特殊类型的循环神经网络 (RNN), 它集成了长短时记忆网络 (LSTM) 的能力和双向网络结构的优势^[19]. 因此 BiLSTM 结合了正向和反向两个方向的数据流, 使得模型不仅能学习到从过去到未来的时序信息, 也能从未来到过去的信息中学习, 这使其在处理需要双向上下文的复杂序列任务时表现出色. 在使用 PID 控制器对永磁同步电机 (PMSM) 的速度进行控制时, 需要综合考虑电机的过去和未来运行状态及性能指标, LSTM 可能无法充分利用未来的信息来调整参数, 而 BiLSTM 通过其双向学习机制, 在预测控制器的响应和优化参数时更为有效. 本研究将 BiLSTM 融合 TD3 的 Actor 网络和 Critic 网络中, 构建了如图 3 所示的 BiLSTM-Actor 和如图 4 所示 BiLSTM-Critic 网络架构. 在 BiLSTM-Actor 网络中设置多个批量标准化层, 每个层都对应一个时间步的输入 S_t, S_{t-1}, S_{t-n} , 用于标准化输入数据, 以确保输入特征在相同的尺度上, 减少内部协变量偏移, 加速训练过程, 并缓解过拟合问题. 标准化的数据进入 BiLSTM 层, 用来处理时间序列数据, 捕捉时间步长之间的长期依赖关系, 之后接入 Dropout 层, 增强模型的泛化能力. 数据经过两层全连接层进行进一步的非线性转换, 得到动作向量 a_t 输出. 对于 BiLSTM-Critic 网

络, 输入状态 s_t 和动作 a_t 合并输入在经过标准化后使用 BiLSTM 对特征进行提取, 最后通过全连接层实现状态-动作到 Q 值的映射.

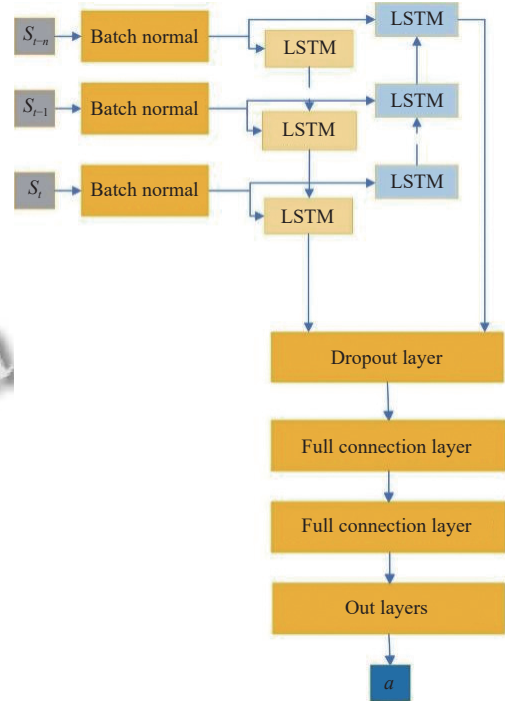


图 3 BiLSTM-Actor

3.4 基于好奇心和熵正则化的联合探索策略

为了促进更强大和多元化的探索, 本文引入了一个基于好奇心和熵正则化的联合探索策略. 这种方法结合了两种策略来鼓励代理更多地探索环境.

(1) 基于好奇心的探索

本文设计了一个额外的神经网络来预测下一个状态, 并根据其预测误差来产生一个“好奇心奖励”^[20]. 这种奖励机制可以鼓励代理更多地探索环境. 内在奖励信号 r_c 计算如下:

$$r_c = \beta \cdot \|s_{t+1} - \hat{s}_{t+1}\|^2 \quad (23)$$

其中, s_{t+1} 是实际的下一个状态, \hat{s}_{t+1} 是通过神经网络预测的下一个状态, β 是一个超参数来控制好奇心奖励的强度.

(2) 熵正则化

本文还引入了熵正则化来鼓励策略保持一个高熵, 这可以避免策略过早地收敛到一个确定性策略. 熵正则化定义如下:

$$E = -\alpha \cdot \sum p(a | s) \cdot \log p(a | s) \quad (24)$$

其中, $p(a|s)$ 是策略产生一个给定动作的概率, α 是一个超参数来控制熵正则化的强度.

在本文策略的实现中, 好奇心奖励和熵正则化项是联合使用的, 它们一起工作来指导代理的探索策略. 具体来说, 将好奇心奖励添加到奖励环境中, 以鼓励代理去探索之前未探索的状态空间. 同时, 还在策略优化

过程中加入了熵正则化项, 来鼓励策略保持一定的探索性. 因此, 本文总奖励函数可以表示为:

$$r_{total} = reward + r_c - \alpha \cdot E \quad (25)$$

通过这种方式, 本文提出的策略不仅可以更好地探索环境, 而且还可以避免过早地收敛到一个确定性策略, 从而在学习过程中保持一定的多样性.

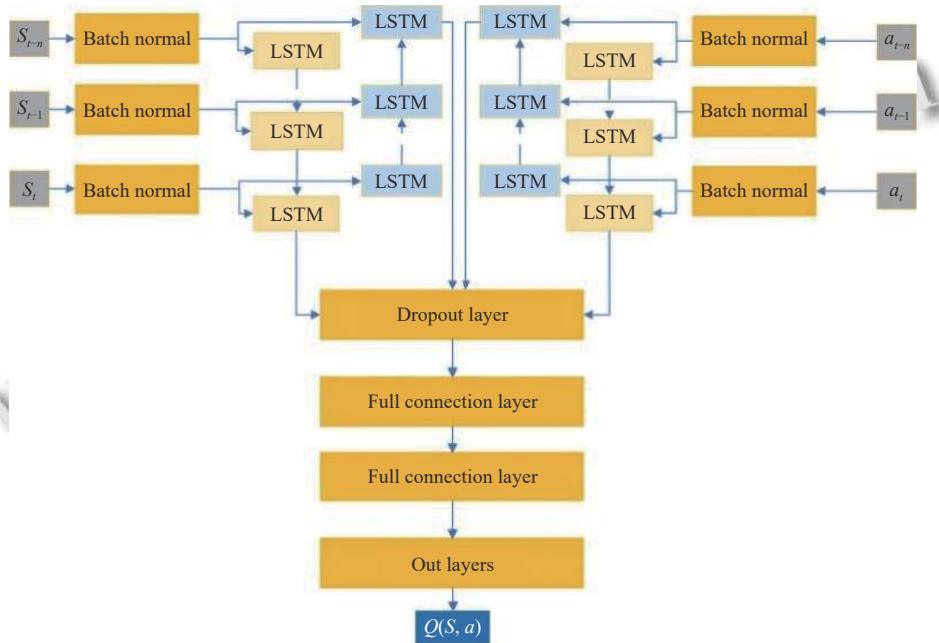


图4 BiLSTM-Critic

4 实验及结果分析

4.1 实验配置

本文基于 PyTorch 平台构建了一个永磁同步电机 PID 控制的仿真环境, 实验中使用了经典的 Ziegler-Nichols (Z-N) 方法、标准的 TD3 算法, 以及本研究新提出的 BiLSTM-TD3-ICE 算法进行比较分析. 本文在不同的参考速度情形下进行了测试, 包括固定速度的空载启动、加入扰动后的空载启动, 以及正弦波和方波参考速度跟踪. 动作空间范围为 $K_P \in [0, 30]$, $K_I \in [0, 5]$. 为了深入评估 TD3 和 BiLSTM-TD3-ICE 在执行目标任务时的性能差异, 本文从奖励分析以及跟踪速度分进行了评估. 算法中涉及的参数如表 2 所示.

4.2 测试实验

4.2.1 奖励分析

本文的分析基于不同时间步的奖励分析来比较两种算法的性能和效率. 对比 TD3 算法与 BiLSTM-TD3-ICE 算法训练的奖励图如图 5 所示.

表 2 BiLSTM-TD3-ICE 算法训练参数

参数名	数值
最大训练轮数	1 000
经验回放池容量	1 000 000
小批量的样本数量	256
软更新率	0.005
奖励折扣因子	0.99
探索噪声	0.1
延迟更新频率	2
熵正则化权重	0.5
好奇心驱动探索权重	0.5

在实验的早期阶段, TD3 算法出现了显著的奖励波动, 这反映了其在进行广泛状态探索时的不稳定性. 该现象是由于 TD3 算法在探索过程中尝试了多种策略以理解环境动态, 导致在某些策略下经历了奖励的下降. 相比之下, BiLSTM-TD3-ICE 算法表现出更加优秀的探索行为和较快的学习速度, 这得益于 BiLSTM 网络的能力在处理时间序列数据方面的优势, 以及结合好奇心激励和熵正则化的探索策略. 随着训练进入

中期和后期阶段, TD3 算法虽然开始展现出一定的学习进步和稳定性, 但相比之下, BiLSTM-TD3-ICE 算法在保持较高奖励水平方面显示了更为显著的稳定性和快速的收敛速度. 总的来说, BiLSTM-TD3-ICE 算法在学习效率、稳定性以及收敛速度方面相较于标准 TD3 算法具有显著的优势, 这些特点使其在处理需要理解和适应长期依赖性信息的复杂环境中更有优势.

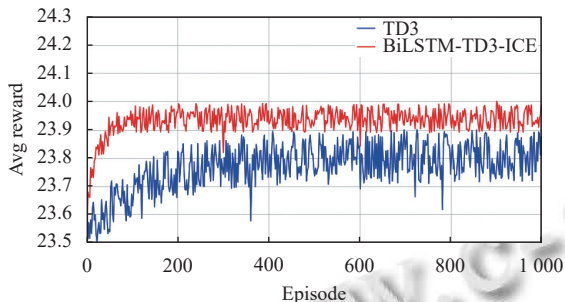


图5 奖励对比图

4.2.2 跟踪速度实验对比分析

在永磁同步电机的速度跟踪实验中, 设定了 4 组实验分别为无干扰下的固定速度 1500 rad/s, 在 0.2 s 处产生扰动的固定速度、转速设置为正弦函数 $100\sin(5\pi t)$ rad/s 以及方波信号. 为此, 采用了经典的 Ziegler-Nichols (Z-N) 方法、传统的 TD3 算法和本文提出的 BiLSTM-TD3-ICE 算法来对 PID 参数进行调整和整定.

对于无干扰下的固定速度的实验如图 6 所示, 与 Z-N 法和传统的 TD3 算法相比, 本文提出的方法能够迅速并准确地跟踪命令信号, 显示出较高的响应速度以及更小的超调量.

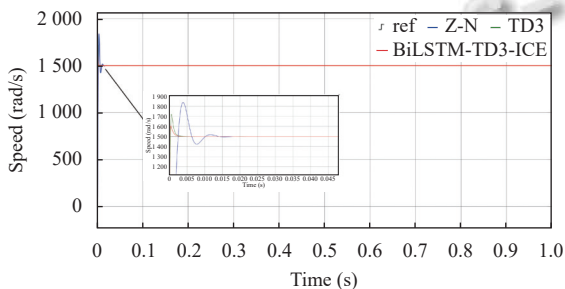


图6 无干扰速度跟踪实验对比图

当在 $t = 0.2$ s 处添加 100 N 的负载扰动, 结果如图 7 所示, BiLSTM-TD3-ICE 算法能够迅速调整 PID 参数以吸收扰动, 使电机速度快速回归到稳定状态, 而传统 TD3 算法和 Z-N 方法则表现出较大的超调和较长的恢复时间.

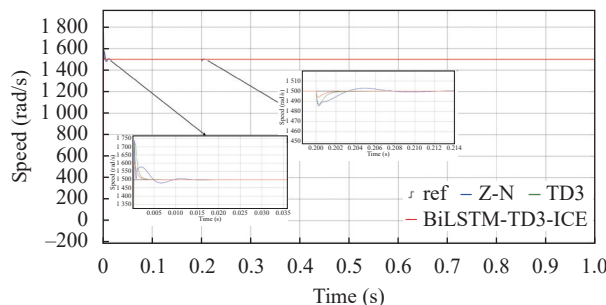


图7 有干扰速度跟踪实验对比图

对于正弦函数参考速度的实验结果如图 8 所示, 我们进一步验证了 Z-N 方法、TD3 和 BiLSTM-TD3-ICE 算法的性能差异. 在正弦函数实验中, 期望的速度呈现出周期性的波动, 这意味着电机需要不断地进行调整以跟随这一变化趋势. 3 种算法都展示了迅速的速度响应. 由于正弦波的周期性特点, 要求电机在高峰和低点之间进行精确的切换. BiLSTM-TD3-ICE 算法对于正弦参考速度的追踪效果显然更为出色. Z-N 方法和 TD3 算法在高峰和低谷的转换过程中存在一些振荡. 相比之下, BiLSTM-TD3-ICE 算法不仅保持了速度的平稳性, 而且在各个转折点上都表现出了较高的精确度, 准确地跟随了期望的速度曲线.

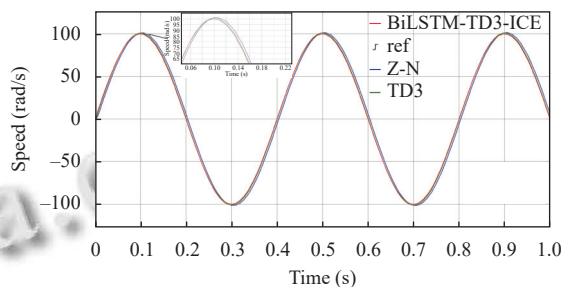


图8 正弦速度跟踪实验对比图

在针对方波信号的实验中, 控制算法需对具有急剧跳变的参考信号进行精确跟踪. 方波信号的特性要求控制系统具备迅速的响应能力以及高度的稳定性, 以便在每次跳变后快速达到新的稳态值, 同时最小化超调和振荡. 实验结果如图 9 显示, BiLSTM-TD3-ICE 算法在处理方波信号时表现出较传统方法更加卓越的性能. 具体来说, 该算法在方波信号的每个上升沿和下降沿处能够实现快速且准确的跟踪, 这得益于 BiLSTM 网络的能力在处理时间序列数据时提供了额外的上下文信息, 有助于算法预测和适应速度的急剧变化. 与 Z-N 方法和传统 TD3 算法相比, BiLSTM-TD3-ICE 在速度

的高峰和低谷处的振荡显著减少,表明其在速度跳变时具有更好的控制稳定性。

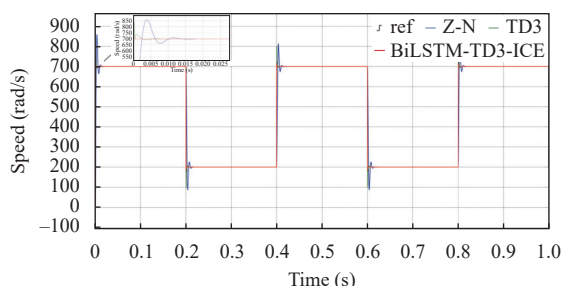


图9 方波速度跟踪实验对比图

本研究通过比较 TD3 算法和 BiLSTM-TD3-ICE 算法在不同速度跟踪场景中的表现,展示了 BiLSTM-TD3-ICE 算法在处理复杂动态系统时的显著优势。在无干扰固定速度跟踪实验中, BiLSTM-TD3-ICE 算法比传统的 Z-N 方法和 TD3 算法显示出更快的响应速度和更低的超调量。当面临突发扰动时, BiLSTM-TD3-ICE 算法能够迅速调整 PID 参数,确保电机速度迅速恢复到稳定状态,而其他算法则在恢复过程中遇到了挑战。特别是在正弦函数和方波信号的跟踪任务中, BiLSTM-TD3-ICE 算法利用 BiLSTM 网络捕捉时间序列的全局特性,实现了高精度的周期性速度跟踪和对急剧跳变信号的准确响应,表现出较少的振荡和更高的控制稳定性。

5 结论

在本文的研究中,针对 PID 控制器优化问题深入探讨,对 TD3 算法进行了多方面的深度优化,本研究通过结合 BiLSTM 网络到 TD3 算法的 Actor 和 Critic 网络中,并引入了内在好奇心和熵正则化策略,从而全面提升了算法的鲁棒性和控制性能。仿真实验结果表明,这些改进使 BiLSTM-TD3-ICE 算法在获得奖励、加速收敛过程以及提升控制精度方面超越了传统 TD3 算法。特别是在复杂的速度跟踪任务中, BiLSTM-TD3-ICE 算法利用 BiLSTM 网络的长短期记忆能力和内在好奇心激励的探索策略,显著提高了对 PMSM 动态的理解和适应能力,进而产生了更为平滑和准确的速度跟踪性能。本文研究成果显著,但仍有优化空间。未来可研究更高级的策略和复杂网络结构来提升系统性能。为增加策略的实用价值,可在更多实际场景中应用并与其他优化技术结合。

参考文献

- 1 Yuan TQ, Wang DZ, Wang XH, *et al.* High-precision servo control of industrial robot driven by PMSM-DTC utilizing composite active vectors. *IEEE Access*, 2019, 7: 7577–7587. [doi: [10.1109/ACCESS.2018.2890539](https://doi.org/10.1109/ACCESS.2018.2890539)]
- 2 Chen H, Wang X, Benbouzid M, *et al.* Improved fractional-order PID controller of a PMSM-based wave compensation system for offshore ship cranes. *Journal of Marine Science and Engineering*, 2022, 10(9): 1238. [doi: [10.3390/jmse10091238](https://doi.org/10.3390/jmse10091238)]
- 3 Chen SC, Hao XH, Gao CZ, *et al.* An effective nontransient active short-circuit method for PMSM in electric vehicles. *IEEE Transactions on Industrial Electronics*, 2023, 70(4): 3571–3580. [doi: [10.1109/TIE.2022.3176315](https://doi.org/10.1109/TIE.2022.3176315)]
- 4 Wang YC, Fang SH, Hu JX. Active disturbance rejection control based on deep reinforcement learning of PMSM for more electric aircraft. *IEEE Transactions on Power Electronics*, 2023, 38(1): 406–416. [doi: [10.1109/TPEL.2022.3206089](https://doi.org/10.1109/TPEL.2022.3206089)]
- 5 Wang YQ, Feng YT, Zhang XG, *et al.* A new reaching law for antidisturbance sliding-mode control of PMSM speed regulation system. *IEEE Transactions on Power Electronics*, 2020, 35(4): 4117–4126. [doi: [10.1109/TPEL.2019.2933613](https://doi.org/10.1109/TPEL.2019.2933613)]
- 6 Ab Ghani MS, Zin NM, Zakaria F, *et al.* Speed control design for permanent magnet synchronous motor drive based on PID-fuzzy controller. *Proceedings of the 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*. IPOH: IEEE, 2022. 170–175.
- 7 Quang NK, Anh NK, Than H, *et al.* Neural network PID controller for PMSM drives. *Proceedings of the 7th National Scientific Conference on Applying New Technology in Green Buildings (ATiGB)*. Da Nang: IEEE, 2022. 146–149.
- 8 Sanatel Ç, Günel GÖ. Long short term memory based self tuning regulator design for nonlinear systems. *Neural Processing Letters*, 2023, 55(3): 3045–3079. [doi: [10.1007/s11063-022-10997-1](https://doi.org/10.1007/s11063-022-10997-1)]
- 9 Xue Z, Wang YY, Li L, *et al.* An adaptive speed control method based on deep reinforcement learning for permanent magnet synchronous motor. In: Jia YM, Zhang WC, Fu YL, *et al.*, eds. *Proceedings of the 2021 Chinese Intelligent Systems Conference: Vol. I*. Singapore: Springer, 2022. 275–286.
- 10 Shi Q, Lam HK, Xiao B, *et al.* Adaptive PID controller based on Q-learning algorithm. *CAAI Transactions on Intelligence Technology*, 2018, 3(4): 235–244. [doi: [10.1049/trit.2018.1007](https://doi.org/10.1049/trit.2018.1007)]

- 11 Liao JR, Liu SY, Wu QH, *et al.* PID control of permanent magnet synchronous motor based on improved Actor-Critic framework. *Journal of Physics: Conference Series*, 2022, 2213: 012005. [doi: [10.1088/1742-6596/2213/1/012005](https://doi.org/10.1088/1742-6596/2213/1/012005)]
- 12 Lu PW, Huang WK, Xiao JL. Speed tracking of brushless DC motor based on deep reinforcement learning and PID. *Proceedings of the 7th International Conference on Condition Monitoring of Machinery in Non-stationary Operations (CMMNO)*. Guangzhou: IEEE, 2021. 130–134.
- 13 Fujimoto S, Van Hoof H, Meger D. Addressing function approximation error in Actor-Critic methods. *Proceedings of the 35th International Conference on Machine Learning*. Stockholm: PMLR, 2018. 1587–1596.
- 14 Shi Q, Lam HK, Xuan CB, *et al.* Adaptive neuro-fuzzy PID controller based on twin delayed deep deterministic policy gradient algorithm. *Neurocomputing*, 2020, 402: 183–194. [doi: [10.1016/j.neucom.2020.03.063](https://doi.org/10.1016/j.neucom.2020.03.063)]
- 15 Ullah K, Guzinski J, Mirza AF. Critical review on robust speed control techniques for permanent magnet synchronous motor (PMSM) speed regulation. *Energies*, 2022, 15(3): 1235. [doi: [10.3390/en15031235](https://doi.org/10.3390/en15031235)]
- 16 Dubey V, Goud H, Sharma PC. Role of PID control techniques in process control system: A review. In: Nanda P, Verma VK, Srivastava S, *et al.*, eds. *Proceedings of the 2021 SSIC Data Engineering for Smart Systems*. Singapore: Springer, 2022. 659–670.
- 17 李茹杨, 彭慧民, 李仁刚, 等. 强化学习算法与应用综述. *计算机系统应用*, 2020, 29(12): 13–25. [doi: [10.15888/j.cnki.csa.007701](https://doi.org/10.15888/j.cnki.csa.007701)]
- 18 Yuan XM, Wang Y, Zhang RC, *et al.* Reinforcement learning control of hydraulic servo system based on TD3 algorithm. *Machines*, 2022, 10(12): 1244. [doi: [10.3390/machines10121244](https://doi.org/10.3390/machines10121244)]
- 19 Siami-Namini S, Tavakoli N, Namin AS. The performance of LSTM and BiLSTM in forecasting time series. *Proceedings of the 2019 IEEE International Conference on Big Data (Big Data)*. Los Angeles: IEEE, 2019. 3285–3292.
- 20 Li J, Shi XX, Li JH, *et al.* Random curiosity-driven exploration in deep reinforcement learning. *Neurocomputing*, 2020, 418: 139–147. [doi: [10.1016/j.neucom.2020.08.024](https://doi.org/10.1016/j.neucom.2020.08.024)]

(校对责编: 张重毅)