伺服系统瞬态优化的模糊自适应深度强化学习方法

魏晓晗[1]，张庆[1,2]，蒋婷婷[1]，梁霖[1,2]
（1. 西安交通大学机械工程学院，710049，西安；
2. 西安交通大学现代设计及转子轴承系统教育部重点实验室，710049，西安）

摘要：为了提高伺服系统瞬态性能，满足工业生产线的高精度加工要求，提出模糊自适应深度强化学习方法，用于永磁同步电机伺服系统的控制性能优化。根据伺服系统瞬态运行的响应快速性和稳定性要求，通过构造奖励函数与 Actor-Critic 网络，在伺服控制器中引入瞬态响应优化环节，面向实时运算需求设计学习率自适应模糊控制器，构建模糊自适应瞬态性能优化方法；建立伺服系统 Simulink 仿真模型，根据实际伺服系统拟合仿真模型，通过瞬态性能仿真获得优化参数，并将优化参数应用于西门子 840D 数控系统。以仿真运算及系统实验对该优化方法进行验证，结果表明，所提出方法使伺服系统调节时间缩短 10％以上，显著提高了系统的瞬态响应速度，且未引入明显超调，验证了方法的可行性。提出的方法具有较强的通用性，为伺服系统智能控制优化提出了新的途径。

关键词：伺服系统；控制优化；瞬态性能；深度强化学习；模糊控制
中图分类号：TP181 **文献标志码**：A
DOI：10.7652/xjtuxb202108009 **文章编号**：0253-987X(2021)08-0068-10

OSID 码

Fuzzy Adaptive Deep Reinforcement Learning Method for Optimizing Transient Performance of Servo System

WEI Xiaohan[1], ZHANG Qing[1,2], JIANG Tingting[1], LIANG Lin[1,2]
(1. School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China; 2. Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: To improve the transient performance of servo systems and satisfy the high-precision processing requirements of industrial production lines, a fuzzy adaptive deep reinforcement learning algorithm is proposed to optimize the control performance of permanent magnet synchronous motor servo system. According to the rapidity and stability requirements of transient response, the reward function and Actor-Critic networks are constructed, the transient optimization link for servo controller is introduced, and the learning rate fuzzy controller for real-time computing requirements is designed. A servo system Simulink simulation model is established and fitted in with the actual servo system, the optimized parameters are obtained via transient performance optimization simulation and then applied to a Siemens 840D computerized numerical control system. The verification of simulation calculation by system experiment shows that the proposed method of transient performance optimization shortens the servo system adjustment time by more than 10％, significantly heightens the system transient response rate

收稿日期：2020-12-02。 作者简介：魏晓晗(1995—)，男，博士生；张庆(通信作者)，男，副教授。 基金项目：国家自然科学基金资助项目(51675405)。
网络出版时间：2021-04-27 网络出版地址：http://kns.cnki.net/kcms/detail/61.1069.T.20210427.1327.006.html

without introducing an obvious overshoot. This method has strong versatility to provide a new way for the intelligent control optimization of the servo system.

Keywords: servo system; control optimization; transient performance; deep reinforcement learning; fuzzy control

伺服系统将控制信号转化为精确的机械运动，是机电装备向智能化、高精度和高效率发展的关键基础，也是生产制造数字化的核心载体。永磁同步电机(PMSM)具有密度大、几何尺寸小、重量轻、响应快、调速范围宽、控制方便等显著优点，已成为伺服系统中将电能向可控机械运动转化的最佳选择之一[1-4]。现代高精度生产线要求伺服系统能够稳定快速地达到控制目标，因此以平滑无超调与快速响应为中心的瞬态响应能力成为评价 PMSM 伺服系统性能的重要指标[5]，也成为电机控制优化的研究热点。

PMSM 伺服系统通过对电机的精确控制实现高精度加工要求，控制优化是目前研究的关键问题。胡强晖等采用全局滑模和变指数趋近率之间的切换控制，有效地消除了永磁同步电机伺服系统的抖振现象，但是由于包含大量的超参数，应用难度较大[6]；吴振顺等提出一种模糊自整定 PID 控制器，对抑制干扰和噪声有一定的效果，但并没有考虑瞬态响应速度，整体性能优化效果有限[7]；蔚永强等采用径向基神经网络对系统进行自适应补偿[8]，有效地改善了直驱式伺服系统中稳态跟踪性能降低的问题，但仅限于仿真研究。由于传统方法在多目标优化能力的限制，现阶段对于伺服系统的研究仍局限于稳态误差消除与鲁棒性提升，然而对于高精度加工而言，提升系统响应速度与抑制超调同等重要。

作为一种适用于多目标优化的方法，深度强化学习方法搜索能力强、收敛速度快，在控制领域展现出了非常高的应用价值。Pane 等结合 Actor-Critic 算法以及 3D 打印机伺服控制系统提出了一种针对伺服系统的稳态误差补偿方法[9]。陈奇石等针对双足机器人稳定行走问题，使用有监督学习中的稀疏在线高斯过程回归方法对强化学习中的函数进行拟合，实现了双足机器人快速、稳定行走的步态规划[10]。Han 提出了一种基于强化学习的在线演化框架来预先检测和修改自动驾驶汽车控制器的不完善决策，兼顾了速度与安全性[11]。由于伺服系统瞬态性能优化中存在建模难度高、仿真运算量大导致优化效率低下的问题，深度强化学习尚未应用于这一领域。

在强化学习中，如何权衡知识的“探索”与经验的“利用”是影响强化学习效率与准确性的重要因素，也是贯穿于强化学习发展史并关系到应用效果的关键性问题。针对这一问题，Schaul 等提出了经验回放方法[12]，即根据重要性对经验池中过往经验样本进行采样。该方法避免了强化学习方法中随机抽取的弊端，提高了算法的收敛速度和准确性，但仅限于 Q-Learning 方式的强化学习算法，无法应用于采用神经网络等非线性策略的近似算法。作为经典的智能控制方法，模糊控制在强耦合复杂模型的控制问题中具有良好的技术优势，通过采用类似优先级采样的策略，有效改善了随机性采样缺陷。因此，通过模糊控制平衡“探索”与“利用”这一对复杂矛盾，是提高强化学习在控制优化领域应用能力的可行思路。

本文以伺服系统瞬态性能提升为目标，提出一种模糊深度强化学习算法，利用异步优势演员-评论家强化学习算法并行运算效率高、收敛速度快、搜索能力强的特点，结合模糊算法在复杂系统控制中鲁棒性强、适应性高的优势，建立基于模糊自适应深度强化学习的伺服系统瞬态性能优化方法。基于 PMSM 伺服系统的仿真计算模型，优化瞬态性能参数，并将模型计算参数应用于实际伺服系统。经验证，该方法有效缩短了永磁同步电机伺服系统调节时间，同时抑制了超调与稳态误差的产生。

1　问题分析

1.1　永磁同步电机伺服系统

永磁同步电动机主要由三相绕组固定定子和永磁体磁极转子组成。当三相交流电源供给定子绕组时，产生旋转磁场。旋转磁场与转子上的磁极相互作用产生电磁转矩，激发转子同步旋转。永磁同步电机原理如图 1 所示。

三相静止坐标系与两相静止坐标系下的定子磁链方程是关于角度 θ 的时变方程，为方便控制器设计，一般在交直轴两相旋转坐标系下对永磁同步电机进行分析。

在交直轴两相旋转坐标系下，伺服电机交直轴电压方程为

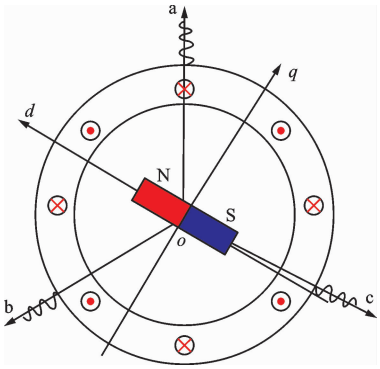

图 1 永磁同步电机原理图

Fig. 1 Principle of permanent magnet synchronous motor

$$\left.\begin{aligned} u_d &= R_s i_d + L_d \frac{\mathrm{d}i_d}{\mathrm{d}t} - \omega L_q i_q \\ u_q &= R_s i_q + L_q \frac{\mathrm{d}i_q}{\mathrm{d}t} + \omega(L_d i_d + \psi_f) \end{aligned}\right\} \tag{1}$$

式中：u_d 为直轴电压；u_q 为交轴电压；R_s 为定子绕组的电阻；i_d 为直轴电流；i_q 为交轴电流；L_d 为直轴等效电感；L_q 为交轴等效电感；ψ_f 为永磁体磁链；ω 为转子角速度。

本文永磁同步电机伺服系统矢量控制原理采用 $i_d=0$ 控制。在 $i_d=0$ 控制方式下，直轴电压转化为

$$u_d = -\omega L_q i_q \tag{2}$$

直轴的电流为 0 时，不再贡献转矩电压。这时电机的所有的电流全部用来产生电磁转矩，只用需要控制交轴电流 i_q 就可以控制电机的转矩，从而实现交直轴电流的静态解耦。

1.2 PMSM 伺服系统瞬态响应

随着 PMSM 伺服系统在工业生产中的广泛应用，日益增长的高精度生产需求要求其动态响应能力快速提升。作为强耦合系统，PMSM 伺服电机与负载系统的匹配质量直接影响电机启、停过程中的机电耦合状态，从而影响电机的响应能力。在 PMSM 伺服系统中，存在电动机驱动、控制参数与力学输出参数、电动机力学输出参数与负载系统等耦合关系，其中电机力学输出与负载耦合方程如下

$$T_e = \frac{3}{2} p(\psi_d i_d - \psi_q i_q) \tag{3}$$

$$J \frac{\mathrm{d}\omega}{\mathrm{d}t} = T_e - T_L - B\omega \tag{4}$$

式中：p 为电机极对数；ψ_d 为直轴等效磁链；ψ_q 为交轴等效磁链；T_e 为电磁转矩；T_L 为负载扭矩；B 为阻尼系数；J 为转动惯量。

在伺服系统瞬态响应优化中，快速性和稳定性是控制的核心目标。从式(3)(4)可以得出，电机响应的快速性由转子角速度 ω 及其微分决定，而响应的稳定性则取决于电机输出转矩、角速度及其微分参数之间的平衡状态。因此，实现伺服电机动态性能优化，需要从伺服系统中的机电耦合关系出发，通过调节电动机的驱动、控制参数，改变电机运动形态，进而影响电动机力学输出参数，从而实现电动机在特定载荷下快速性与稳定性的平衡，使电机与负载系统达到最佳匹配状态。

随着相关研究者对 PMSM 的深入研究，许多对于闭环控制参数的优化控制方法如滑模变结构控制、模糊控制和自适应控制等[13-16]，位置的响应速度与控制精度得到了显著的提升。传统参数整定寻优算法无法兼顾包括响应速度和控制精度在内的多个控制目标，或对于复杂耦合问题搜索能力欠佳，因此难以解决伺服系统中复杂耦合关系下的驱动、控制系统参数优化问题。寻求一种适用于多目标、多参数、复杂耦合问题的控制优化方法，是解决永磁同步电机伺服系统瞬态性能优化问题的关键。

作为一种适用于复杂控制、决策问题的机器学习算法，深度强化学习算法[17-18]在解决多目标、多参数问题中具有收敛效率高，稳定性强的优势，使之成为解决 PMSM 伺服系统瞬态性能优化问题的可行方案。在本文针对伺服系统瞬态性能优化问题，对强化学习算法进行模糊自适应改进，提高收敛速度以满足实际应用中时效性的要求。

2 模糊自适应深度强化学习方法

2.1 强化学习原理

强化学习是一种数据驱动的在线启发式机器学习方法，是行为主体从环境到行为映射的学习[19]。作为一种仿生算法[20]，在强化学习中，智能体通过与环境交互实现学习过程。智能体从环境中以状态信号的形式接收状态信息，并根据策略函数 $\pi(s,a)$ 采取相应的动作。策略函数 π 往往是随机的，通过与奖励函数结合实现对每一个可能的状态转换的评估，并根据概率选取动作。在算法的每一步执行动作时，环境将受到动作影响，根据状态转移函数与奖励函数转移到下一个状态 s'，并得到瞬时奖励 r_t。根据状态 S_t 计算得到神经网络输出 V_t 作为本状态价值函数的估计，由新状态 S_{t+1} 更新输出 V_{t+1}。经过 N 步后完成一个回合的运算，并得到回报函数 G_t。根据 Bellman 方程，价值函数定义为 t 时刻状态 S_t 能获得的回报期望，其表达式如下

$$V^{\pi}(s)=E_{\pi}\left\{\sum_{k=0}^{\infty}\gamma^{k}r_{t+k+1}\mid S_t=s\right\}=E_{\pi}\{r_t+\gamma V^{\pi}(S_{t+1})\mid S_t=s\}\tag{5}$$

式中：γ 为折扣因子。

最优累计期望由 V^* 表示，将 $G_t=r_{t+1}+\gamma r_{t+2}+\cdots=\sum_{k=0}^{\infty}\gamma^{k}r_{t+k+1}$ 代入式(5)，得到最优价值函数为

$$V^*(s)=\max_{a\in A(s)}E_{\pi}(G_t\mid S_t=s,A_t=a)=\max_{a\in A(s)}\sum_{s',r}p(s',r\mid s,a)[r+\gamma V^*(s')]\tag{6}$$

式中：$A(s)$ 为动作空间；$p(s',r|s,a)=\Pr\{S_{t+1}=s',r_{t+1}=r|S_t=s,A_t=a\}$，即在状态 $S_t=s$ 下实施动作 $A_t=a$，得到下一个状态 $S_{t+1}=s'$ 与奖励 $r_{t+1}=r$ 的概率。

在强化学习中，迭代运算的目标在于训练得到最优控制策略 $\pi(s,a)$ 最大化累计奖励 G。以价值函数最优化为方向进行迭代，即可实现累计奖励最大化。

2.2　异步优势演员评论家(A3C)算法

异步优势演员-评论家(A3C)算法是一种基于并行计算和 Actor-Critic 算法的深度强化学习算法。其中 Actor-Critic 算法包括 Actor 和 Critic 两个网络，Actor 是一个以策略为基础的网络，通过奖惩信息调节不同状态下采取各种动作的概率；Critic 是一个以值为基础的学习网络，可以计算每一步的奖惩值。二者相结合，在不断迭代中，得到每一个状态下选择每一动作的合理概率。在 A3C 算法中，创建多个并行的环境，每个并行环境同时运行 Actor 与 Critic 网络，让多个拥有副结构的智能体同时在并行环境上更新主结构中的参数。并行运算中的智能体互不干扰，而全局网络的参数更新则通过每个本地网络分别上传的运算后的更新梯度实现，这种更新方式降低了算法中经验之间的相关性，因此收敛性显著提高。

在 A3C 算法中，每一个步长 t 中，Actor 网络将会根据状态 S_t 和策略 $\pi(a_t|S_t;\theta'_{\mathrm{a}})$ 选择动作 a_t（θ'_{a} 为 Actor 网络参数），这使得环境转换到下一个状态 S_{t+1}，并得到一个瞬时奖励 r_t。Critic 网络估计本状态的价值函数，将瞬时奖励与下一个状态的价值估计之差作为本状态价值函数的训练目标，因此 Critic 网络的时序优势函数为

$$f_{\mathrm{adv}}(s,t)=r_t+\gamma V(S_{t+1};\theta'_{\mathrm{v}})-V(S_t;\theta'_{\mathrm{v}})\tag{7}$$

式中：θ'_{v} 为 Critic 网络参数。

根据 Bellman 方程，优势函数 $f_{\mathrm{adv}}(s,t)$ 体现了价值函数 $V(S_t;\theta'_{\mathrm{v}})$ 的优化程度。

策略函数 $\pi(a_t|S_t;\theta')$ 的损失函数为

$$L_{\mathrm{a}}=\nabla_{\theta'}\log\pi(a_i\mid S_t;\theta')(R-V(S_t;\theta'_{\mathrm{v}}))+c\nabla_{\theta'}H(\pi(S_t;\theta))\tag{8}$$

式中：$c\nabla_{\theta'}H(\pi(S_t;\theta))$ 为策略 $\pi(a_t|S_t;\theta')$ 的熵项；价值函数的损失函数为

$$L_{\mathrm{c}}=R-V(S_t;\theta'_{\mathrm{v}})\tag{9}$$

对参数进行梯度更新时，Actor 网络和 Critic 网络采取的策略分别为

$$\mathrm{d}\theta=\nabla_{\theta'}\log\pi(a_i\mid S_t;\theta')(R-V(S_t;\theta'_{\mathrm{v}}))+c\nabla_{\theta'}H(\pi(S_t;\theta))\tag{10}$$

$$\mathrm{d}\theta_{\mathrm{v}}=\partial(R-V(S_t;\theta'_{\mathrm{v}}))^2/\partial\theta\tag{11}$$

根据该更新策略迭代后输出最终的动作结果。

2.3　模糊自适应深度强化学习算法

在工业应用中，伺服系统由于工况复杂多变以及在寿命周期中的状态变化等问题，优化控制时需要考虑实时性需求，对强化学习算法的计算效率提出了更高的要求，因此必须对强化学习中的“探索”与“利用”机制进行综合平衡。参考生物对于经验的记忆程度的区别，采用的改进策略为：区分经验的优先级重要性，并根据经验的优先级对经验区别记忆。

针对包括 Q-Learning 在内的所有强化学习算法，区分经验重要性并实现记忆区分的一种可行思路为：通过经验中动作得到的奖励来区分经验重要性，并根据重要性调整学习率，从而实现经验记忆的优先级更新。这种方法借鉴大脑对于不同记忆的深刻程度的区别，对于“非同寻常”的经验印象更加深刻，而对于“司空见惯”的经验则记忆程度较低，能够保证生物在进行正常学习的同时保持对异常现象的警觉性，从而使生物获得的知识更全面。在强化学习中，动作奖励表示智能体所采取动作的优劣，而学习率影响每条经验的更新程度，即记忆程度。相应地，对于获得收益或惩罚较大的经验采用较大的学习率，对于其他一般经验采用较小的学习率，以期在提高收敛速度的同时保证准确性。

为实现该策略，需要寻找一种行之有效的手段，以得到经验优先级与记忆深度之间的映射。模糊控制算法[21]自 1965 年提出以来，广泛应用于工业控制领域，模糊隶属度作为多因素影响下的综合决策评价函数，在系统性能优化方面体现出强大的优势。因此，在强化学习过程中建立模糊控制器，解决经验优先级与记忆深度的平衡问题。针对奖励偏差及其变化率建立论域$[-e,e]$，通过隶属度函数运算得到输入值的模糊语言，在模糊控制规则的条件语句中搜索对应学习率控制量的模糊语言，最后对控制量

模糊语言进行清晰化，从而将推理得到的学习率控制量转化为学习率输出。算法流程如图 2 所示。

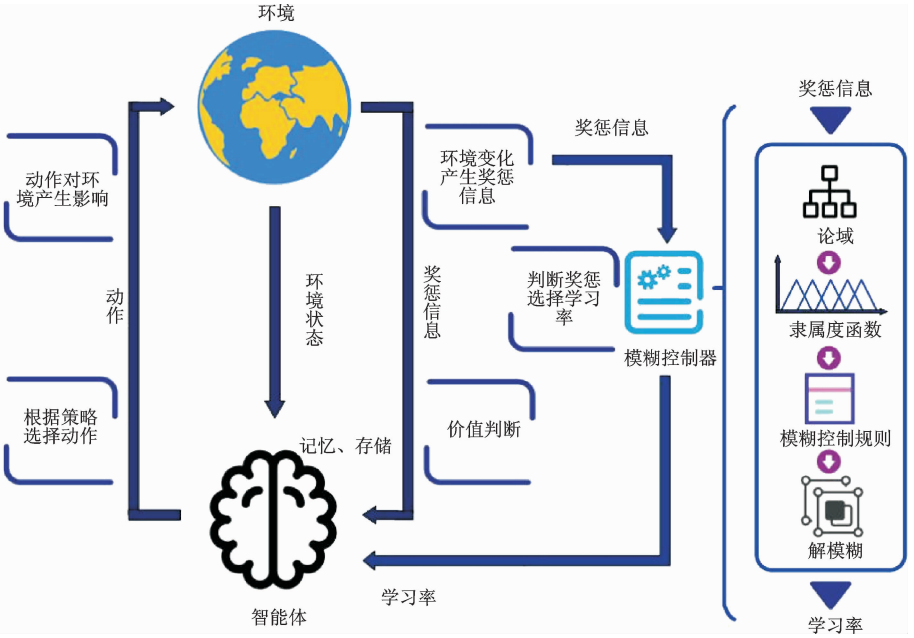

图 2 模糊自适应深度强化学习算法流程

Fig. 2 Fuzzy adaptive reinforcement learning algorithm flow

3 瞬态性能优化方法

在伺服系统瞬态性能优化中，需要同时考虑响应的快速性与稳定性两个方面，使系统响应速度得以提升，同时避免演化为欠阻尼系统产生超调量。针对该问题，基于深度强化学习的永磁同步电机伺服系统瞬态性能优化方法在 PID 环节中设计补偿环节，根据缩短调节时间、抑制超调量的多个控制目标建立评价指标，利用评价指标设计 A3C 算法奖励函数，并根据控制补偿环节确定 A3C 算法动作参数，确定最优补偿参数。

如图 3 所示，在机器学习工作站中建立伺服系统仿真模型，并根据实际伺服系统参数及运行结果拟合仿真模型，减小模型与系统偏差；建立 A3C 强化学习算法全局及本地 Actor-Critic 网络，根据缩短调节时间、抑制超调量的控制目标建立强化学习价值函数；在 A3C 强化学习算法中引入学习率模糊控制器，根据经验优先级调整学习率，提高运算效率。

3.1 价值函数设计

瞬态性能综合优化的目的在于缩短调节时间、

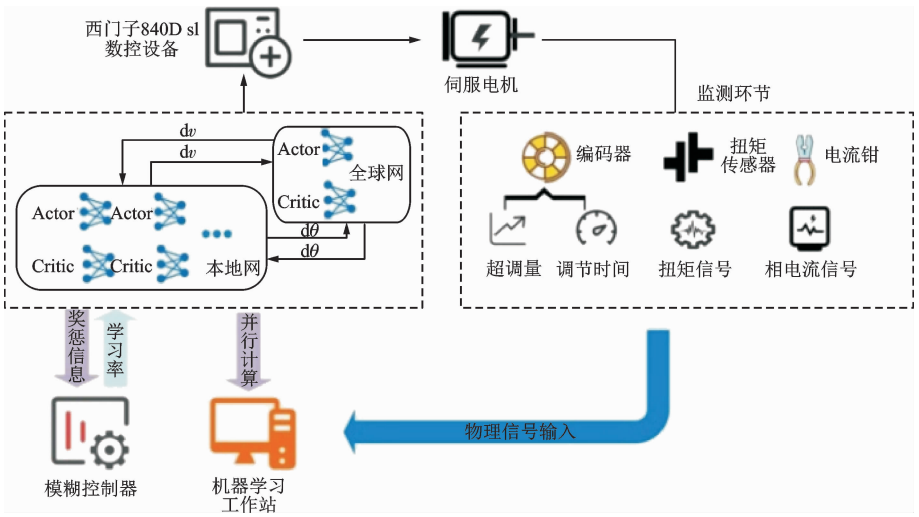

图 3 基于模糊自适应深度强化学习算法的伺服系统瞬态性能优化

Fig. 3 Transient performance optimization of servo system based on fuzzy adaptive deep reinforcement learning algorithm

抑制超调量，同时保证正常工作效率。因此，建立伺服系统调节时间 t_s、超调量 $\sigma\%$、效率指标 η（电流与扭矩有效值之比）3 个瞬态响应性能指标作为算法评价指标。设置评价指标向量即状态向量

$$\boldsymbol{S}_t = \{\sigma\%, t_s, \eta_s\} \tag{12}$$

其中

$$\sigma\% = \frac{c_{tp} - c_\infty}{c_\infty} \times 100\% \tag{13}$$

$$t_s = t' \tag{14}$$

$$\eta_s = \frac{T_{rms}}{I_{rms}} \tag{15}$$

式中：c_{tp} 为伺服系统位置时域响应最大偏离值；c_∞ 为伺服系统位置时域响应终值；t' 为伺服系统位置时域响应稳定至终值 98% 所用的时间；T_{rms} 为伺服系统扭矩时域响应有效值；I_{rms} 为伺服系统电流时域响应有效值。

对状态向量中各元素进行归一化处理，得到归一化后的状态向量 $\boldsymbol{S}_t^a$。引入重要性矩阵

$$\boldsymbol{I}_{imp} = [i_1 \quad i_2 \quad i_3]^T \tag{16}$$

式中：i_1、i_2、i_3 为重要性因子。设置奖励函数和线性补偿函数为

$$\boldsymbol{R}(s_t, a) = (\boldsymbol{R}_{ref} - \boldsymbol{S}_t^a)\boldsymbol{I}_{imp} \tag{17}$$

$$\theta_c = K\left(-\frac{1}{t_d}t + 1\right) \tag{18}$$

式中：$\boldsymbol{R}_{ref}$ 为奖励函数参考向量；t_d 为补偿截止时间。选择补偿放大增益 K 作为算法输出动作，作为 Actor 网络的输出，s_t 和 K 作为 Critic 网络的输入。

3.2　Actor-Critic 网络设计

由于径向基神经网络逼近精度高、结构简单、线性拟合，以及便于进行微分运算的特点，在 Actor 与 Critic 网络的参数化设计中采用该网络。

根据输入状态及动作分别设计 Actor-Critic 算法中 Actor 与 Critic 参数化网络参数 $\phi(s)$，在[0,1]中等间隔设置神经网络中心，随机配置初始 Actor 网络权值参数 θ 与 Critic 网络权值参数 ω_i。径向基神经网络形式如下

$$y_n = \sum_{i=1}^{n} \omega_i \sum_k \mathrm{e}^{\frac{\| x_k - c_i \|^2}{2\sigma^2}} \tag{19}$$

式中：c_i 为径向基中心；n 为隐含层网络节点数；k 为输入层节点个数。

对网络参数 ω_i 进行偏导运算

$$\frac{\partial y_n}{\partial \omega_i} = \sum_{i=1}^{n} \sum_k \mathrm{e}^{\frac{\| x_k - c_i \|^2}{2\sigma^2}} \tag{20}$$

使用径向基网络设计的 Actor-Critic 网络，在进行梯度更新时，能够快速地通过微分运算得到新参数，从而提高计算效率。运行 A3C 算法迭代至收敛，即可得到伺服系统补偿参数。

3.3　模糊控制器设计

模糊控制器的设计包括论域划分、确定模糊集合、隶属度函数求解、模糊规则推理和解模糊等步骤。定义给定的奖励偏差为本回合本步收益与前 N 回合平均收益值之差，根据运算经验，设定奖励偏差 e、变化率 e_c 及输出值学习率 r_1 的基本论域。输入量与输出量所取的模糊子集的论域都为{−6，−5，−4，−3，−2，−1，0，1，2，3，4，5，6}，对应的模糊语言为{负大、负中、负小、零、正小、正中、正大}。隶属度函数选用三角形函数的形式，三角形隶属度函数分布曲线如图 4 所示。

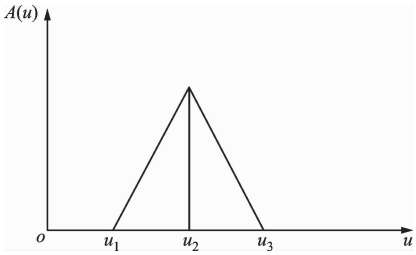

图 4　三角形隶属度函数分布曲线

Fig. 4　Triangular membership function distribution curve

三角形隶属度函数表达式如下

$$a(u) = \begin{cases} (u - u_{i-1})/(u_i - u_{i-1}), u \in [u_1, u_2] \\ (u_{i+1} - u)/(u_{i+1} - u_i), u \in [u_2, u_3] \\ 0, u \in (-\infty, u_1) \cup (u_3, +\infty) \end{cases} \tag{21}$$

本文基于三角形隶属度函数 $a(u)$ 设计输入与输出量隶属度函数 $A(u)$ 如图 5 所示，图中{NB，NM，NS，ZO，PS，PM，PB}对应模糊语言{负大、负中、负小、零、正小、正中、正大}。

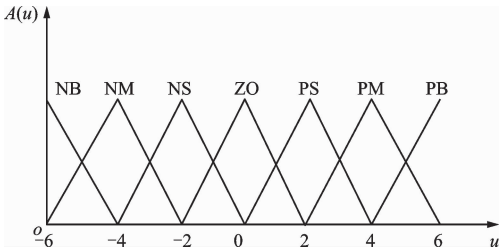

图 5　输入量与输出量隶属度函数

Fig. 5　Membership function of input and output

根据隶属度函数可以得到特定奖励偏差 e 及其变化率 e_c 对应的模糊语言。在进行模糊推理时，需

要考虑在不同情况下奖励偏差 e 及其变化率 e_c 的模糊语言之间的逻辑关系，并根据逻辑关系确定模糊推理结果。因此，根据学习率整定的工程经验建立模糊推理规则表，如表 1 所示。

表 1 输出值模糊规则表

Table 1 Output value fuzzy rule table

e	e_c						
	NB	NM	NS	ZO	PS	PM	PB
NB	PB	PB	PB	PB	PM	ZO	ZO
NM	PB	PB	PB	PB	PM	ZO	ZO
NS	PM	PM	PM	PM	ZO	NS	NS
ZO	NM	NM	NS	ZO	NS	NM	NM
PS	PS	PS	ZO	NM	NM	NM	NM
PM	ZO	ZO	NM	NB	NB	NB	NB
PB	ZO	ZO	NM	NB	NB	NB	NB

表 1 中任意一条规则都可表示为 IF e is A_{i1} and e_c is A_{i2}，THEN r_1 is C_i 的形式。通过上述的逻辑及条件推理，总共可以将模糊规则表总结为 49 条模糊控制条件语句。

为得到精确输出值，需要根据学习率的模糊语言，结合隶属度函数，通过面积中心法对输出值进行反模糊化。面积中心法表达式如下

$$U=\frac{\sum_i \mu_i A(u_i)}{\sum_i A(u_i)} \tag{22}$$

通过反模糊化，得出学习率 r_1 的精确输出 U。

4 仿真实验

为提高运算效率，同时确保系统安全性，本文利用仿真运算得到优化参数，在仿真模型中验证优化效果，继而应用于实际伺服系统。选择西门子 840D sl 数控系统为实验对象，包括驱动器、控制器、执行器与检测器件 4 个部分，控制参数如表 2 所示。

表 2 西门子 840D sl 数控系统控制参数

Table 2 Siemens 840D sl control parameters

控制环	比例增益	微(积)分时间常数
位置环	1	0.000 1
速度环	0.452	0.166 6
电流环	20.038	3.333×10^{-6}

驱动电机为西门子 1FK7083-2AF71-1AA0 永磁同步电机，参数如表 3 所示。

表 3 永磁同步电机设计参数

Table 3 Permanent magnet synchronous motor design parameters

参数	数值	参数	数值
扭矩常量/($N\cdot m\cdot A^{-1}$)	1.58	磁链/Wb	0.017
电压常数/($V\cdot(r\cdot min^{-1})^{-1}$)	0.024 7	转动惯量/($kg\cdot m^2$)	0.002 6
旋转电感/mH	7.0	黏滞摩擦系数/($N\cdot m\cdot s^{-1}$)	0.000 388
绕组电阻/Ω	0.38	极对数	8
电枢电感/H	0.008 35	额定扭矩/($N\cdot m$)	10.5

4.1 伺服系统仿真模型

在 Simulink 仿真环境中，对 PMSM 伺服系统建立如图 6 所示的仿真模型，包括 PID 控制器、被控对象、执行、检测、比较、补偿 6 个环节。该伺服系统仿真模型采用位置控制方式，主要由通用电桥模块、永磁同步电机模块、空间矢量脉宽调制模块、PID 控制器、坐标变换模块 5 个模块组成，模拟 840D sl 数控系统驱动进给轴运动过程。

根据伺服系统电机硬件参数与控制参数配置仿真模型，使模型逼近真实伺服系统。参照表 2 及表 3，配置图 7 所示模型，经过仿真运算，得到相同行程下伺服系统运动轨迹。以 600 mm 行程为例，轨迹曲线如图 7 所示。

实际伺服系统±5%的调节时间为 7.06 s，而仿真伺服系统±5%的调节时间约为 6.47 s，两者差距为 8.49%；实际伺服系统上升时间约为 5.32 s，仿真模型上升时间约为 5.02 s，两者之间差距为 5.97%。本文所用到的仿真模型为简化模型，而西门子 840D sl 系统中存在复杂的驱动参数与通道控制参数，这导致伺服系统实际运行相应曲线与仿真曲线形状上略有差异，但从瞬态响应性能方面可以认为二者性能相近，符合实验要求。

4.2 结果及分析

进行仿真运算，加入模糊控制器后算法在大约第 53 回合收敛，最大奖励值为 13.58，而未加入模糊控制器时算法大约在 152 回合收敛，最大奖励值为 13.53。这表明，在加入学习率模糊控制器后，收敛准确性未受影响，而收敛速度得到提升。迭代过程中奖励函数变化与未加入模糊控制器变化表现对比如图 8 所示。

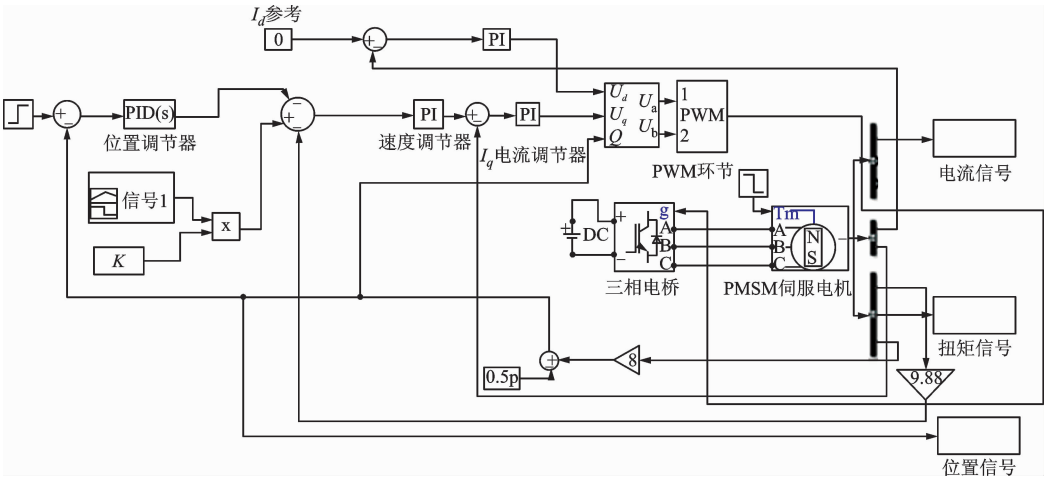

图 6　永磁同步电机伺服系统 Simulink 仿真模型

Fig. 6　Simulink simulation model of permanent magnet synchronous motor servo system

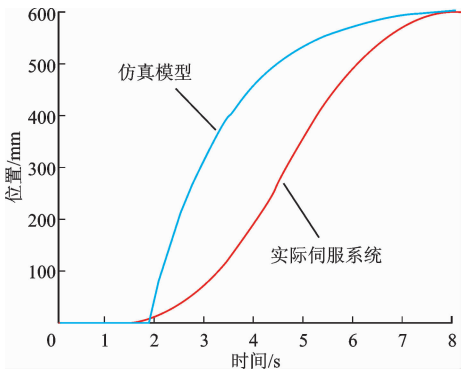

图 7　仿真模型与伺服系统响应曲线对比

Fig. 7　Comparison of simulation model with servo system response

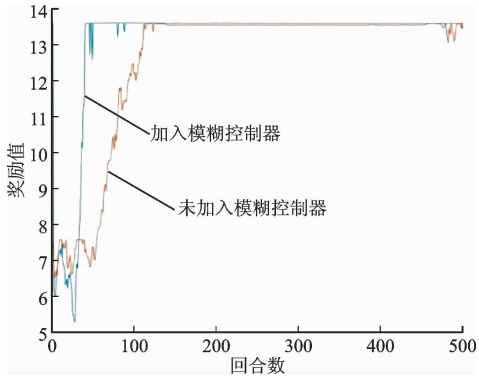

图 8　优化仿真奖励函数变化过程对比

Fig. 8　Comparison of reward function change process in simulation

经过运算得到最优补偿增益 $K=53.564\ 5$，加入补偿增益得到仿真补偿运行结果与原始运行结果对比如图 9 所示。根据响应曲线可以得出仿真前后主要响应指标如表 4 所示，补偿前的伺服系统调节时间为 5.02 s，超调量为 0；补偿后的调节时间为 4.13 s，超调量约为 0.3%。在未引入明显超调的前提下，调节时间缩短 17.72%。

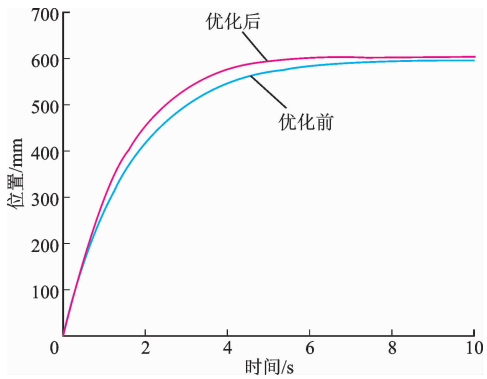

图 9　仿真模型优化效果对比

Fig. 9　Comparison of simulation model optimization effects

表 4　仿真系统优化瞬态响应指标对比

Table 4　Comparison of transient response indexes for simulation system optimization

状态	调节时间/s	超调量/%	效率指标
优化前	5.02	0	0.265 3
优化后	4.13	0.3	0.285 7

由仿真结果可知，模糊自适应深度强化学习伺服系统瞬态优化方法可以成功提高系统响应速度，伺服系统的调节时间和效率都有不同程度的提升。加入模糊控制器，在未影响算法准确性的同时，大幅提升了运算效率。

4.3　系统应用

实际系统如图 10 所示，包括 SINAMICS S120 驱动系统、SIMATIC S7-300 PLC 控制系统与 HMI

人机交互系统，负载装置选用三菱 ZKB-XN 磁粉制动器。

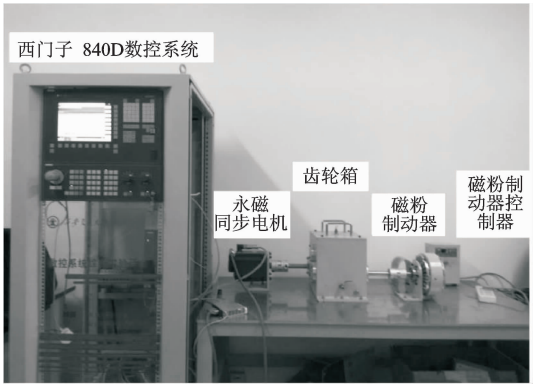

图 10 西门子 840D sl 数控系统试验台

Fig. 10 Siemens 840D sl CNC system test bench

将仿真实验得到的优化参数输入实际数控系统中，运行数控伺服系统后得到补偿前后伺服系统响应对比如图 11 所示。补偿前的调节时间为 5.32 s，超调量为 0；补偿后的调节时间为 4.75 s，超调量为 0。优化后在没有引入超调的前提下，系统位置调节时间缩短 10.71%。优化前后响应指标对比如表 5 所示。

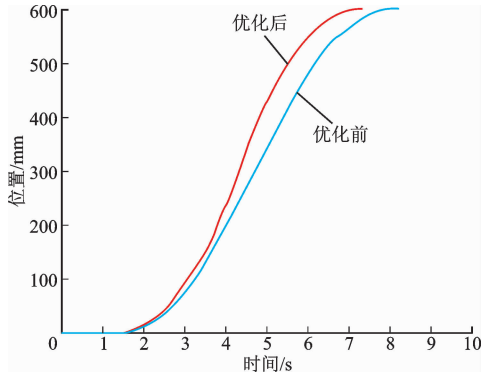

图 11 伺服系统实验台优化响应曲线

Fig. 11 Effect of servo system optimization

表 5 伺服系统实验台优化瞬态响应指标对比

Table 5 Comparison of transient response indexes for optimization of servo system test bench

状态	调节时间/s	超调量	效率指标
优化前	5.32	0	0.242 1
优化后	4.75	0	0.248 3

由实验可以得出，模糊自适应深度强化学习方法针对缩短调节时间与抑制超调的多目标优化问题，明显提升了实际伺服系统瞬态性能。在引入瞬态性能并行优化运算得到的优化参数之后，伺服系统的调节时间有较大程度的缩短。与此同时，补偿优化并没有引入额外稳态误差，对超调也有一定的抑制作用，验证了该方法在提升伺服系统瞬态性能中的有效性。

5 结 论

通过改进深度强化学习算法，建立伺服系统瞬态性能优化方法，针对伺服系统特点，即运行存在时滞性、欠阻尼系统存在超调的问题，设计奖励函数，利用 A3C 优化算法，得到优化补偿函数。结合伺服系统仿真模型进行优化仿真，并在西门子数控系统中验证了优化效果。结果表明，在没有引入超调量的前提下对伺服系统的响应时间产生了明显的提升。由于本文方法具有易实施的特点，在优化控制领域中将具有良好的发展潜力。

参考文献：

[1] 鲁文其，胡育文，梁骄雁，等. 永磁同步电机伺服系统抗扰动自适应控制 [J]. 中国电机工程学报，2011，31(3)：75-81.
LU Wenqi，HU Yuwen，LIANG Jiaoyan，et al. Anti-disturbance adaptive control for permanent magnet synchronous motor servo system [J]. Proceedings of the CSEE，2011，31(3)：75-81.

[2] 刘海财，李光军，李树胜. 基于改进型滑模观测器 PMSM 转子位置估计 [J]. 微特电机，2017，45(11)：23-25，29.
LIU Haicai，LI Guangjun，LI Shusheng. Rotor position estimation method for permanent magnet synchronous motor based on the modified sliding mode observer [J]. Small & Special Electrical Machines，2017，45(11)：23-25，29.

[3] 邢向上，姜新建. 飞轮储能系统电机及其控制器概述 [J]. 储能科学与技术，2015，4(2)：147-152.
XING Xiangshang，JIANG Xinjian. Introduction to motors and controllers of flywheel energy storage systems [J]. Energy Storage Science and Technology，2015，4(2)：147-152.

[4] CALNETIX Technologies. Vycon direct connect kinetic energy storage systems [EB/OL]. [2020-11-20]. http://www.Calnetix.com.

[5] ZHANG Qing，TAN Luyao，XU Guanghua. Evaluating transient performance of servo mechanisms by analysing stator current of PMSM [J]. Mechanical Systems and Signal Processing，2018，101：535-548.

[6] 胡强晖，胡勤丰. 全局滑模控制在永磁同步电机位置伺服中的应用 [J]. 中国电机工程学报，2011，31

(18)：61-66.
HU Qianghui，HU Qinfeng. Global sliding mode control for permanent magnet synchronous motor servo system [J]. Proceedings of the CSEE，2011，31(18)：61-66.

[7] 吴振顺，姚建均，岳东海. 模糊自整定 PID 控制器的设计及其应用 [J]. 哈尔滨工业大学学报，2004，36(11)：1578-1580.
WU Zhenshun，YAO Jianjun，YUE Donghai. A self-tuning fuzy PID controller and its application [J]. Journal of Harbin Institute of Technology，2004，36(11)：1578-1580.

[8] 蔚永强，郭宏，谢占明. 直驱式伺服系统的神经网络自适应滑模控制 [J]. 电工技术学报，2009，24(3)：74-79，85.
WEI Yongqiang，GUO Hong，XIE Zhanming. Neural network adaptive sliding mode control for direct-drive servo system [J]. Transactions of China Electrotechnical Society，2009，24(3)：74-79，85.

[9] PANE Y P，NAGESHRAO S P，KOBER J，et al. Reinforcement learning based compensation methods for robot manipulators [J]. Engineering Applications of Artificial Intelligence，2019，78：236-247.

[10] 陈奇石. 强化学习在仿人机器人行走稳定控制上的研究及实现 [D]. 广州：华南理工大学，2016：41-56.

[11] HAN T，NAGESHRAO S，FILEV D P，et al. An online evolving framework for advancing reinforcement-learning based automated vehicle control [J]. IFAC-PapersOnLine，2020，53：8118-8123.

[12] HORGAN D，QUAN J，BUDDEN D，et al. Distributed prioritized experience replay [EB/OL]. [2020-11-20]. https：//arxiv. org/abs/1803. 00933.

[13] 张肖平，于金飞，崔英英，等. 考虑铁损的永磁同步电动机有限时间模糊动态面控制 [J]. 机械制造与自动化，2021，50(1)：49-53，56.
ZHANG Xiaoping，YU Jinfei，CUI Yingying，et al. Finite-time fuzzy dynamic surface control of PMSM with iron losses [J]. Machine Building & Automation，2021，50(1)：49-53，56.

[14] 陈玄，李祥飞，周杨. 基于新型滑模扰动观测器的永磁同步电机控制 [J]. 电机与控制应用，2020，47(3)：23-27.
CHEN Xuan，LI Xiangfei，ZHOU Yang. Permanent magnet synchronous motor control based on new sliding mode disturbance observer [J]. Electric Machines & Control Application，2020，47(3)：23-27.

[15] 潘玉成，林鹤之，陈小利，等. 基于模糊 RBF 神经网络的 PID 控制方法及应用 [J]. 机械制造与自动化，2019，48(3)：215-219.
PAN Yucheng，LIN Hezhi，CHEN Xiaoli，et al. PID control method based on fuzzy-RBF neural network and its applicaiton [J]. Machine Building & Automation，2019，48(3)：215-219.

[16] 赵珊珊，李晓庆，纪志成. 基于模型参考模糊自适应控制的永磁同步电机控制器设计 [J]. 电机与控制应用，2005，32(8)：20-25.
ZHAO Shanshan，LI Xiaoqing，JI Zhicheng. A new design of speed controller of PMSM based on model reference fuzzy adaptive control method [J]. Electric Machines & Control Application，2005，32(8)：20-25.

[17] 张腾，张小栋，张英杰，等. 引入深度强化学习思想的脑-机协作精密操控方法 [J]. 西安交通大学学报，2021，55(2)：1-9.
ZHANG Teng，ZHANG Xiaodong，ZHANG Yingjie，et al. A precise control method for brain-computer cooperation with deep reinforcement learning [J]. Journal of Xi'an Jiaotong University，2021，55(2)：1-9.

[18] 赵星宇，丁世飞. 深度强化学习研究综述 [J]. 计算机科学，2018，45(7)：1-6.
ZHAO Xingyu，DING Shifei. Research on deep reinforcement learning [J]. Computer Science，2018，45(7)：1-6.

[19] 夏金，孙宏波，孙立民. 基于强化学习的生产再决策问题 [J]. 计算机集成制造系统，2019，25(11)：2935-2942.
XIA Jin，SUN Hongbo，SUN Limin. Reinforcement learning for production reschedule [J]. Computer Integrated Manufacturing Systems，2019，25(11)：2935-2942.

[20] DABNEY W，KURTH-NELSON Z，UCHIDA N，et al. A distributional code for value in dopamine-based reinforcement learning [J]. Nature，2020，577(7792)：671-675.

[21] 郑飞，吴钦木. 基于模糊 PI 的永磁同步电机矢量控制 [J]. 计算机与现代化，2021(1)：7-11.
ZHENG Fei，WU Qinmu. Vector control of permanent magnet synchronous motor based on fuzzy PI [J]. Computer and Modernization，2021(1)：7-11.

（编辑　杜秀杰）