文章编号:1674-2974(2024)12-0176-10 DOI:10.16339/j.cnki.hdxbzkb.2024294

基于强化学习的舞台多轴同步系统预测维护策略研究

李炜[1,2],王洁莹[1],毛海杰[1,2†]
(1. 兰州理工大学 电气工程与信息工程学院,甘肃 兰州 730050;
2. 甘肃省工业过程先进控制重点实验室,甘肃 兰州 730050)

摘 要:针对舞台多轴同步系统因执行器退化导致无法满足控制任务时限要求,现有维护策略难以达优的问题,提出一种基于强化学习的舞台多轴同步系统预测维护策略.首先将强化学习以串级方式引入,构建具有寿命预测与自主维护能力,能以不同采样率分而治之的控制架构;其次,聚焦介入维护策略及多源不确定性对执行器退化过程的影响,基于卡尔曼(Kalman)滤波、期望最大化和固定间隔平滑等算法,通过对执行器退化状态的实时感知、估计及退化模型的自适应更新,确保多轴同步系统剩余寿命预测精度;结合系统期望工作时限与剩余寿命预测的偏差、执行器实时退化状态等构建Q-learning算法的目标函数,通过不断试错对维护控制量做出最优调整,以获得最大的寿命延长奖励,从而实现了舞台多轴同步系统智能优化维护.通过舞台多轴同步系统仿真实验验证了所提方法的有效性,提高了系统维护效能.

关键词:舞台多轴同步;执行器退化;剩余寿命预测;强化学习;预测维护
中图分类号:TP273 **文献标志码**:A

Research on Predictive Maintenance Strategy for Stage Multi-axis Synchronous System Based on Reinforcement Learning

LI Wei[1,2],WANG Jieying[1],MAO Haijie[1,2†]
(1. College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China;
2. Key Laboratory of Gansu Advanced Control for Industrial Processes, Lanzhou 730050, China)

Abstract: Aiming at the problem that the performance of the stage multi-axis synchronous system cannot meet the time limit of the control task due to the degradation of the actuators, and the existing maintenance strategy is difficult to reach the optimization, this paper proposes a reinforcement learning-based predictive maintenance strategy for the stage multi-axis synchronous system. Firstly, reinforcement learning is introduced in a cascaded manner, and constructing a control architecture with capabilities for lifespan prediction and autonomous maintenance, which operates with different sampling rates. Secondly, focusing on the intervening maintenance strategy and the influence of multi-source uncertainty on the actuator degradation process, based on the algorithms of Kalman filtering, Expectation-Maximum, and Rauch-Tung-Striebel smoothing, by the real-time perception and estimation of actuator degradation state, and a daptive update of degradation model, the prediction accuracy of the

* **收稿日期**:2023-10-21
基金项目:国家自然科学基金资助项目(62063017), National Natural Science Foundation of China(62063017)
作者简介:李炜(1963—),女,陕西周至人,兰州理工大学教授
† 通信联系人,E-mail: gsmaohj@163.com

remaining life of the multi-axis synchronous system is ensured. Combined with the real-time perception, deviation of remaining life prediction, and the actuator degradation state, the objective function of a Q-learning algorithm is constructed. The optimal adjustment of maintenance control is carried out through trials and errors to obtain the maximum life extension reward and realize intelligent optimization maintenance of the stage multi-axis synchronous system. Finally, the effectiveness of the proposed method is verified by simulation experiments of the stage multi-axis synchronous control system, improving the system maintenance efficiency.

Key words: stage multi-axis synchronous system; actuator degradation; residual life prediction; reinforcement learning; predictive maintenance

舞台造型的多层次变化决定了其展现效果及表现力，而这种变化又依赖于舞台控制系统. 多轴同步控制因其具有更大负载驱动能力、高效灵活运动方式、可协调系统各轴间的运动等优势，已广泛应用于舞台演绎领域[1-2]. 随着剧场演出频次的增加，在长期运行过程中，舞台系统中各种部件不可避免会受疲劳磨损、环境腐蚀、负载变化以及扰动等影响，导致系统性能和健康状态逐渐退化，进而出现安全隐患甚至引发安全事故[3]. 由于舞台多轴同步系统的特殊性，人与设备常常需要动态交互，且有较为严格的工作时限要求，因此，通过适时适当的维护策略，确保其安全可靠运行显得尤为重要.

近年来，随着对重大装备安全性要求的提高和信息、计算等技术的迅猛发展，预测性维护(predictive maintenance, PdM)作为一种新型的维护策略应运而生[4-5]. PdM 利用系统剩余使用寿命(remaining useful life, RUL)信息，通过实施更具前瞻性的维护策略，在提升系统安全性的同时提高维护效率，已在多个领域受到关注[6-8]. 其中针对具有工作时限要求的控制系统，PdM 研究则是基于系统 RUL 的实时预测信息，主要借助于线性二次调节器(linear quadratic regulator, LQR)[9]、模型预测控制(model predictive control, MPC)[10-11]等控制策略，在系统失效前通过在线参数调节等措施实施自主维护，以延长其使用寿命. 其中 MPC 也因特有的鲁棒性和适应性更受青睐. 文献[12-13]分别针对升降舞台、四旋翼无人机等多轴系统的健康自主维护问题，开展了初探性研究，其基于系统 RUL 预测，通过对动态矩阵控制(dynamic matrix control, DMC)中约束矩阵的调节，减缓执行器退化，有效地延长了系统的使用寿命. 目前基于 MPC 维护的调节方式均是围绕约束矩阵值调节展开，主要包括固定值和自适应变值两种. 前者需人工反复试凑，且缺少与退化状态关联；后者虽通过构造调整函数与退化状态关联，但函数的参数依旧需要人工试凑. 因而这种依赖人工经验的维护方式，势必难达最优.

随着人工智能技术的迅猛发展，强化学习(reinforcement learning, RL)方法在解决大规模复杂决策问题方面展现出巨大的潜力. RL 通过与未知环境的交互学习，获取反馈信号并不断调整策略，以逐步提高智能体的性能[14-15]. RL 亦可处理连续状态和动作空间，如控制问题中的连续控制变量. RL 在描述动态函数方面的可处理性以及表征各种系统动态特征的灵活性，使其能够在基于模型和无模型的环境中学习，并与环境交互得到最优控制策略. 文献[16]针对单回路控制系统，首次基于 RL 提出了一种健康感知的控制策略，通过对系统真实退化下预测的 RUL 与期望的 RUL 做差，利用 RL 试错学习获得最优控制策略，有效延长了系统寿命. 因此，如何将 RL 应用于舞台多轴同步系统的健康自主维护，规避 MPC 维护方法对人工经验的依赖，获得多回路耦合系统更有效的维护策略，无疑对提高系统安全性是极有意义的.

要实现舞台多轴同步系统的健康 PdM，前提是根据部件退化数据准确预测系统的 RUL. 目前有关独立设备或部件级设备的 RUL 预测方法已较为成熟，根据建模方式不同，寿命预测方法可分为基于机理模型的方法、基于数据驱动的方法以及二者相融合的方法[12]. 随着传感与数据采集技术的发展，基于数据驱动的方法逐渐成为主流. 依据对数据的处理方式不同，其又分为机器学习的方法和随机过程的方法等. 其中，基于机器学习的 RUL 预测方法

虽然能够获取具体的预测数值，但需要大量的退化数据进行模型训练，对于退化存在随机性、时变性及退化数据缺失的舞台多轴同步系统并不适用. 文献[17]对大数据背景下随机退化设备的RUL技术进行了系统的综述. 基于随机过程的方法关键是选择合适的随机模型，其中Wiener过程因能描述非单调退化过程，并可获得RUL预测的解析解等优点，成为建模方法的首选[18-22]. 然而现存成果在退化建模时，尽管对退化过程的时变不确定性、传感器测量噪声影响、退化数据缺失等问题已有或多或少的考虑，但维护策略介入后对系统退化的影响尚未涉及，而且以独立设备和部件为多，控制系统也以单回路为主. 因此，考虑介入维护策略及多源不确定性的综合影响，如何基于Wiener过程，通过对执行器退化过程精准建模，提高舞台多轴同步系统RUL预测精度，是有效实现PdM的关键.

此外，无论是单回路还是多回路系统，考虑到系统部件退化通常都会经历长期缓慢的变化过程，而系统状态则因需要变化较快，因此，健康维护策略的施加频率也就无须与系统状态监测、控制的频率相同，相反短周期监控与长周期维护干预不仅有利于节约计算与存储资源，而且对于系统控制本身和维护的时效性也更加符合实际情况. 因此，针对舞台多轴同步系统，如何通过层次结构的区分和控制策略的选择，构建一个基于多采样率的健康感知反馈调节的智能体系架构，是实现预测性维护(PdM)的基础.

鉴于此，本文旨在针对舞台多轴同步系统的预测维护需求，开展如下创新性研究工作：

首先以串级方式将RL引入舞台多轴同步系统，构建了在维护层与控制层，能以不同采样率分而治之的智能PdM体系架构；其次，考虑多源不确定及介入维护策略后对执行器退化模型的影响，利用Wiener过程对其退化建模，采用Kalman滤波、期望最大化(Expectation-Maximum，EM)及固定间隔(Rauch- Tung-Striebel，RTS)平滑算法对退化状态估计、模型参数自适应更新，以获取系统更精准的RUL；再者，利用系统寿命偏差、执行器实时退化状态等构建Q-learning算法的目标函数，在不断试错中对维护控制量做出最优调整，获得最大的寿命延长奖励，以实现舞台多轴同步系统健康的更有效维护.

1 基于RL的舞台多轴同步系统PdM框架构建

在实际舞台多轴同步系统中，PID算法作为广泛且有效应用的控制策略，已取得工程界的高度认可[12]. 并联方式下的偏差耦合控制因具有同步性能好、结构稳定可靠等优点[23]，也是舞台同步系统常用的控制结构. 因此，本文底层(控制层)依旧采用偏差耦合控制结构和PID控制器，仍以串级的思想，在其上层(维护层)增添一个PdM模块，如图1所示(仅考虑单轴退化的情形).

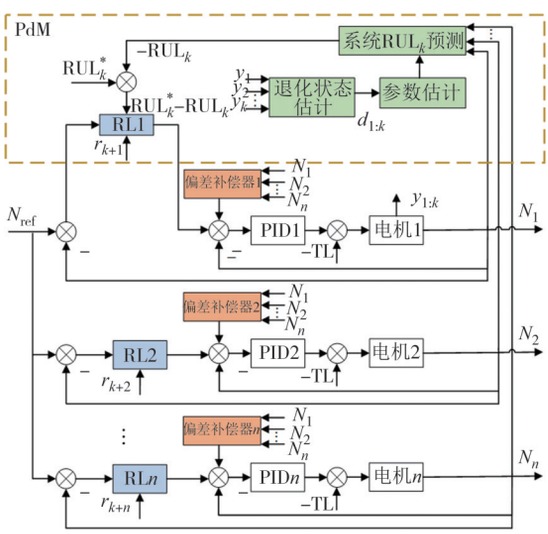


图1 基于RL的舞台多轴同步系统PdM框架

Fig.1 PdM architecture for stage multi-axis synchronous systems based on RL

图1中：N_{ref}为期望位置设定值；RLn、PIDn分别表示第n个RL控制器和第n个PID控制器；TL为扰动输入；N_n为各轴位置的实际输出量；y_k表示已有退化监测数据；d_k表示真实退化状态；RUL$_k$、RUL$^*_k$分别为预测的系统使用寿命与期望的系统使用寿命；r_{k+n}表示奖励函数.

本文与文献[12]不同之处在于以下三点：

1)RL控制器取代了DMC控制器. RL的优势在于无需先验知识和模型，而受文献[16]启发将系统使用寿命误差(RUL* - RUL)植入目标函数，则能通过在线自主试错学习，获取减小或消除系统使用寿命误差(RUL*-RUL)的最优控制策略，提高预测维护的精准性，使系统使用寿命得以有效延长，达到PdM的目的. RL维护策略的输入为RUL*与RUL的差值，其中RUL*为期望的系统使用寿命，RUL为根据实际

执行器退化状态、系统输出等感知信息预测的系统使用寿命.

2)不同层级以不同采样率分而治之工作.在基于RL的舞台多轴同步系统健康自主维护PdM双层架构中,根据控制层与维护层的属性与需求不同,为兼顾系统控制性能和维护效率,分别为控制层、维护层设定较高、较低的采样率,尤其是基于RL的长周期维护层,可为早期退化获得初始最优维护策略提供较充分的学习时长,为后期RL快速获得最优维护策略积累经验,从而在确保维护实时性和节约计算资源的同时,提升跟踪调节和自主维护性能.

3)退化模型参数可自适应更新.综合考虑退化数据缺失、传感器测量噪声、维护策略干预、退化过程时变等多源不确定性等因素,借鉴文献[23]思想,基于Kalman滤波、EM及RTS平滑算法,实现退化状态、Wiener过程退化模型参数的自适应估计和更新,确保舞台多轴系统RUL预测精度.

2 RUL的自适应预测

2.1 执行器退化过程

电机作为舞台系统的执行器,在长期运转过程中,因疲劳磨损、环境腐蚀、负载变化等影响,会出现轴偏心、绝缘层老化、磁性退化、电刷磨损等现象,致使电机执行能力下降.考虑舞台多轴系统在运行过程中,退化数据严重缺失,测量噪声和误差无法避免,个体退化因系统结构、工作场景的差异,尤其是在线自主维护促使的退化过程演变,均会加剧执行器退化的时变和非线性特征.兼顾电机的退化属性和上述不确定因素,为建立更为精准的执行器退化模型,采用基于Wiener过程的自适应退化建模方法[23]无疑是可行的,其退化过程可描述为:

$$Y(t)=D(t)+e_{\varepsilon}(t)=D(0)+\int_0^t\mu(\tau;\ \theta_k)\mathrm{d}\tau+\sigma_{B_k}B(t)+e_{\varepsilon}(t) \tag{1}$$

式中:$\{Y(t),t\geqslant 0\}$表示t时刻测量数据;$D(t)$为执行器的退化值;$D(0)$为执行器的初始退化值,通常取0;$B(t)$表示标准布朗运动;σ_{B_k}为扩散系数;$e_{\varepsilon}(t)\sim N(0,\sigma_{\eta_k}^2)$,表示与$B(t)$相互独立的随机测量误差;$\mu(\tau;\ \theta_k)$是漂移系数,根据电机实际退化轨迹,选用幂函数模型$\mu(\tau;\ \theta_k)=a_kb_kt^{b_k-1}$描述其非线性特征;$\theta_k=[a_k,b_k,\sigma_{\eta_k},\sigma_{B_k}]$表示时变参数集合.

2.2 实际退化状态估计与模型参数自适应更新

如何从式(1)中获取真实的退化信息$D(t)$,并确保退化模型参数的实时更新,是准确预测舞台多轴同步系统RUL的基础.

令$y_k=Y(t_k)$表示t_k监测时刻得到的执行器退化监测数据,则$Y_{1:k}=(y_1,y_2,\cdots,y_k)$为系统初始运行时刻至当前监测时刻$t_k$的已有退化监测数据,其真实退化状态为$D_{1:k}=(d_1,d_2,\cdots,d_k)$.其离散形式的执行器退化模型可表示为式(2):

$$\begin{cases}d_k=d_{k-1}+a_k(k\Delta t)^{b_k}-a_k\left[(k-1)\Delta t\right]^{b_k}+v_k\\ y_k=d_k+e_{\varepsilon k}\end{cases} \tag{2}$$

式中:$\Delta t=t_k-t_{k-1}$表示离散化步长;$e_{\varepsilon k}$表示t_k监测时刻的e_{ε}具体值;$v_k=\sigma_{B_k}\left[B(t_k)-B(t_{k-1})\right]$且$e_{\varepsilon k}$是独立同分布的随机噪声序列,故有$v_k\sim N\left(0,\sigma_{B_k}^2(t_k-t_{k-1})\right)$.

由于退化模型对时间呈现非线性特征,但对状态呈现线性特征,故可采用Kalman滤波技术基于监测数据$Y_{1:k}$对真实退化状态$D_{1:k}$进行估计.定义隐含退化量d_k滤波的期望和方差分别为$\hat{d}_{k|k}=E\left(d_k|Y_{1:k},\theta_k\right)$和$P_{k|k}=\mathrm{var}\left(d_k|Y_{1:k},\theta_k\right)$.

具体状态估计过程如下:

1)状态估计:

$$\hat{d}_{k|k-1}=\hat{d}_{k-1|k-1}+\theta_k\Delta t\sqrt{b^2-4ac} \tag{3}$$

$$P_{k|k-1}=P_{k-1|k-1}+\sigma_{B_k}^2\Delta t \tag{4}$$

$$K(k)=P_{k|k-1}(P_{k|k-1}+\sigma^2)^{-1} \tag{5}$$

$$\hat{d}_{k|k}=\hat{d}_{k|k-1}+K(k)(y_k-\hat{d}_{k|k-1}) \tag{6}$$

2)方差更新:

$$P_{k|k}=\left[1-K(k)\right]P_{k|k-1} \tag{7}$$

由于退化过程的时变性,固定的模型参数难以对其精确描述,因此需对每一时刻模型参数进行更新.由于状态估计中包含了模型未知参数,导致极大似然估计算法不再适用,故而采用对于缺失或隐含数据具有特殊优势的EM算法.

模型未知参数θ_k的自适应更新过程如下:

1)利用历史退化数据,得到模型初始参数θ_0;

2)利用对数似然函数对第i步迭代的参数估计值$\hat{\theta}_k^i$求条件期望;

$$\ell\left(\theta_k,\hat{\theta}_k^i\right)=E_{\hat{\theta}_k^i}\left\{L\left(D_{1:k},Y_{1:k}\middle|\theta_k\right)\middle|Y_{1:k}\right\} \tag{8}$$

3)最大化$\ell(\theta_k,\hat{\theta}_k^i)$;

$$\hat{\theta}_k^{i+1}=\arg\max_{\theta_k^i}\ell(\theta_k,\hat{\theta}_k^i) \tag{9}$$

迭代步骤2)和步骤3)直到满足收敛判据终止.

4)求解隐含变量的条件期望值:

通过RTS算法计算$E_{d_{0:k}|Y_k,\hat{\theta}_k^i}(d_k)$, $E_{d_{0:k}|Y_k,\hat{\theta}_k^i}(d_k^2)$, $E_{d_{0:k}|Y_k,\hat{\theta}_k^i}(d_kd_{k-1})$. 算法RTS流程和公式详见文献[23].

通过Kalman滤波、EM及RTS平滑等算法,便可实现退化模型参数的自适应估计和更新,减少退化数据缺失、个体差异、维护干预的影响,提高退化建模的准确性,确保所建模型能更好地契合真实的退化过程.

2.3 舞台多轴同步系统剩余寿命的自适应预测

考虑舞台多轴同步系统中多个执行器同时退化的复杂性,本文先从单一执行器退化下舞台多轴同步系统的剩余寿命预测入手.这里仍以文献[12]同步误差ω作为系统性能指标,结合首达时间定义系统RUL.对多轴同步控制系统的同步误差ω,其数学描述如式(10)所示:

$$\omega=\max\left(\left|N_i-N_j\right|\right),i,j=1,2,\cdots,n且i\neq j \tag{10}$$

式中:N_i,N_j表示各轴的实际位置输出量.

定义同步误差$\{\omega(t),t\geqslant 0\}$首次超出失效阈值$\omega_{th}$的时刻即为多轴同步系统的失效时刻,舞台多轴同步控制系统的寿命为:

$$T_{\mathrm{f}}=\inf\{D(t)|D(t)>D_{\mathrm{th}}\} \tag{11}$$

其中,inf为下确界算子.

退化量阈值D_{th}定义为:

$$D_{\mathrm{th}}=\inf\{D(t)|\omega>\omega_{\max}\} \tag{12}$$

则舞台多轴同步系统的剩余寿命RUL为:

$$\mathrm{RUL}(t)=\inf\{T_{\mathrm{f}}|T_{\mathrm{f}}>t,D(t)>D_{\mathrm{th}}\}-t \tag{13}$$

基于退化模型公式(1),可求解得到RUL的概率分布解析表达式[23]为:

$$\begin{aligned}&f_L(l_k|\theta_k,D_{1:k})\approx\\&\left\{D_{\mathrm{th}}-\lambda(l_k;\theta_k)+l_k\mu(l_k+t_k;\theta_k)-\frac{P_{k|k}\left[\omega-\lambda(l_k;\theta_k)\right]+\hat{d}_{k|k}\sigma_{B_k}^2l_k}{P_{k|k}+\sigma_{B_k}^2l_k}\right\}\times\\&\frac{1}{\sqrt{2\pi l_k^2(P_{k|k}+\sigma_{B_k}^2l_k)}}\exp\left\{-\frac{\left[D_{\mathrm{th}}-\lambda(l_k;\theta_k)-\hat{d}_{k|k}\right]^2}{2\left(P_{k|k}+\sigma_{B_k}^2l_k\right)}\right\}\end{aligned} \tag{14}$$

式中:$\lambda(l_k;\theta_k)=a_k(l_k+t_k)^{b_k}-a_kt_k^{b_k}$;$\mu(l_k+t_k;\theta_k)=a_kb_k(l_k+t_k)^{b-1}$.

可以看出,由于时变退化过程的模型参数会在每一时刻实时更新为$\hat{\theta}_k$,舞台多轴同步系统RUL也随之实现了自适应预测.

3 基于RL的舞台多轴同步系统健康自主维护策略

3.1 Q-learning算法

RL机理是通过与环境的交互进行试错来更新策略,以逐渐接近最优策略.常见的RL算法有Q-learning、Sarsa、Deep Q-Network(DQN)等[13]. Q-learning算法简单易实现且收敛性好;Sarsa适用于连续动作空间但收敛性较差;DQN学习能力强但复杂且易陷入局部最优解.综合考虑本文选用Q-learning算法.

Q-learning是一种基础的RL算法,用于解决马尔可夫决策问题.在Q-learning中,智能体根据Q表进行动作选择,并使用映射方式存储状态和动作之间的对应关系.每个状态-动作对都有一个$Q_k(d_k,u_k)$值,其中,d_k表示t_k时刻的状态,u_k表示t_k时刻的动作.算法开始时,Q表被初始化.智能体通过随机选择动作并观察环境给予的奖励来更新Q表中的值.当智能体再次遇到相同的状态时,它会根据$Q_k(d_k,u_k)$选择具有最高奖励的动作.为了保持对新策略的探索,智能体采用贪婪策略选择动作,即以ε的概率选择Q值最大的动作,而以$1-\varepsilon$概率随机选取动作,其更新公式如下:

$$u_k=\begin{cases}\arg\max Q(d_k,u_k), & 1-\varepsilon\\ \forall\mathrm{random}(U), & \varepsilon\end{cases} \tag{15}$$

式中:ε是取值范围为[0,1]的随机变量.当动作u_k被执行以后,系统将从当前状态d_k进入下一状态d_{k+1},并收到来自环境的奖励.此时,Q表中的映射对$Q_{k+1}(d_k,u_k)$将按照式(16)的方式进行更新:

$$\begin{aligned}&Q_{k+1}(d_k,u_k)\leftarrow Q_k(d_k,u_k)+\\&\alpha\left[r_{k+1}+\gamma\max_{u'}Q_k\left(d_{k+1},u'\right)-Q_k(d_k,u_k)\right]\end{aligned} \tag{16}$$

式中:学习速率$\alpha\in(0,1]$;折扣系数$\gamma\in[0,1]$,γ决定了未来收益的现值.

一般情况下奖励函数与状态d_k和动作u_k有关,满足下式:

$$r_{k+1}\left(d_k,u_k\right)\leftarrow\left(d_k^{\mathrm{T}}\boldsymbol{A}d_k+u_k^{\mathrm{T}}\boldsymbol{B}u_k\right) \tag{17}$$

式中:$\boldsymbol{A}$,$\boldsymbol{B}$分别为半正定矩阵.

在Q-learning算法中,奖励函数的形式与系统

当前状态和智能体采取的动作密切相关.这个更新过程会不断重复,直到智能体学会在每个状态下采取适当的策略,并获得相对稳定的最优决策.

3.2 基于Q-learning算法的舞台多轴同步系统健康自主维护策略

Q-learning算法是基于无模型的强化学习方法,它通过不断更新Q值函数进行学习.在Q-learning中,奖励函数的设计直接影响到Q值函数的更新和学习效果.因此,根据需求设计一个合适的奖励函数,对于Q-learning算法的性能和收敛性非常重要.对于舞台多轴同步系统,由图1可知,Q-learning算法位于上层,负责维护控制策略,其目的在于,当执行器退化导致系统剩余寿命RUL_k不满足工作时限预期RUL_k^*时,通过自主学习获取最优策略,以尽可能延长系统使用寿命,得到最大回报.由于退化过程缓慢,所以不必采用内环PID的快速更新频率,这样既可以节约计算与储存资源,也给Q-learning算法寻优赢得了时间.为确保维护的实效性,维护层(外环)也采用相对控制层(内环)缓慢的更新频率.

因此,基于Q-learning算法原理,构建舞台多轴同步系统维护策略的奖励函数r_{k+1}为:

$$\begin{aligned} r_{k+1} = -\frac{1}{2}\Big[& d_k^{\mathrm{T}}\boldsymbol{S}d_k + \left(\mathrm{RUL}_k^* - \mathrm{RUL}_k\right)^{\mathrm{T}} \times \\ & \boldsymbol{P}\left(\mathrm{RUL}_k^* - \mathrm{RUL}_k\right) + u_k^T\boldsymbol{R}u_k\Big] = \\ & \rho\left(d_k, \mathrm{RUL}_k, u_k\right) \end{aligned} \tag{18}$$

式中既包含了当前执行器的退化状态d_k,也包含了剩余寿命偏差$\left(\mathrm{RUL}_k^* - \mathrm{RUL}_k\right)$.$\boldsymbol{S}$,$\boldsymbol{R}$,$\boldsymbol{P}$为半正定常数矩阵,奖励函数本质上为二次函数.不难看出要使奖励r_{k+1}最大,需根据当前的退化状态d_k,寻求最佳的控制量u_k,使剩余寿命偏差最小.

类似式(16)引入折扣系数γ,可以得到累计奖励$R^h\left(d_k, \mathrm{RUL}_k, u_k\right)$由系统退化状态$d_k$和$\mathrm{RUL}_k$构成,且遵循控制策略$u_k = h(d_k)$:

$$\begin{aligned} & R^h\left(d_k, \mathrm{RUL}_k, u_k\right) = \sum_{i=k}^{\infty}\gamma^{i-k}r_{i+1} = \\ & \sum_{i=k}^{\infty}\gamma^{i-k}\rho\left(d_i, \mathrm{RUL}_i, h\left(d_i\right)\right) \end{aligned} \tag{19}$$

其状态动作价值函数Q函数为:

$$\begin{aligned} Q^h\left(d_k, \mathrm{RUL}_k, u_k\right) = & \rho\left(d_k, \mathrm{RUL}_k, u_k\right) + \\ & \gamma R^h\left(d_{k+1}, \mathrm{RUL}_{k+1}, h\left(d_{k+1}\right)\right) \end{aligned} \tag{20}$$

Q函数的Bellman方程为:

$$\begin{aligned} Q^h\left(d_k, \mathrm{RUL}_k, u_k\right) = & \rho\left(d_k, \mathrm{RUL}_k, u_k\right) + \\ & \gamma Q^h\left(d_{k+1}, \mathrm{RUL}_{k+1}, h\left(d_{k+1}\right)\right) \end{aligned} \tag{21}$$

当式(20)、式(21)是控制策略u_k的最优解$u_k^* = h^*(d_k)$的最大值Q^*时,式(21)可表示为:

$$Q^*\left(d_k, \mathrm{RUL}_k, u_k\right) = \max_{u\in U} Q^h\left(d_k, \mathrm{RUL}_k, u_k\right) \tag{22}$$

$$\begin{aligned} Q^*\left(d_k, u_k\right) = & \rho\left(d_k, u_k\right) + \\ & \gamma \max_{u'} Q^*\left(f_s\left(d_k, u_k\right), u_k'\right) \end{aligned} \tag{23}$$

最优控制h^*:

$$h^*\left(d_k\right) = \arg\max_{u\in U} Q^*\left(d_k, \mathrm{RUL}_k, u_k\right) \tag{24}$$

依据式(16),基于Q-learning算法的舞台多轴同步系统健康维护策略的Q值可进一步表示为:

$$\begin{aligned} Q_{k+1}\left(d_k, \mathrm{RUL}_k, u_k\right) = & Q_k\left(d_k, \mathrm{RUL}_k, u_k\right) + \\ & \alpha_k\Bigg[\underbrace{\underbrace{r_{k+1} + \gamma\max_{u'}Q_k\left(d_{k+1}, \mathrm{RUL}_{k+1}, u'\right)}_{\text{updated estimate}} - \underbrace{Q_k\left(d_k, u_k\right)}_{\text{current estimate}}}_{\text{temporal difference}}\Bigg] \end{aligned} \tag{25}$$

控制动作的策略依据式(15)进行选取.

3.3 基于Q-learning舞台多轴同步系统PdM算法流程与步骤

综合上述内容,基于Q-learning舞台多轴系统PdM算法流程如图2所示.主要包括:执行器退化量的估计、退化模型参数的自适应更新、系统剩余寿命的实时预测以及基于Q-learning算法的自主维护.特别是Q-learning算法的引入,基于退化状态和寿命误差,可在不断试错中对控制量做出最优调整,并获得最大的寿命延长奖励.周而复始,直到达到维护目标或寿命终止.

4 仿真实验与结果分析

4.1 实验描述

为验证本文所提方法的有效性,以两轴舞台升降系统为案例.考虑升降舞台通常采用变频器和三相异步电机作为驱动装置,为简化建模过程,采用矢量变换法将变频器和异步电机等效为直流电机[12],其传递函数为:

$$G(s) = \frac{1/C_{\mathrm{e}}}{T_{\mathrm{m}}T_{\mathrm{l}}s^2 + T_{\mathrm{m}}s + 1} \tag{26a}$$

$$T_{\mathrm{m}} = 2JR/C_{\mathrm{e}}C_{\mathrm{m}} \tag{26b}$$

$$T_{\mathrm{l}} = L/R \tag{26c}$$

式中:C_{e}为反电势系数;J为电机转子的电动惯量;R为定子绕组电阻;L为绕组等效电感;C_{m}电磁转矩常数;T_{m}为机电时间常数;T_{l}为电磁时间常数;与文献[12]相同,可得$T_{\mathrm{m}} = 0.075\ \mathrm{s}$,$T_{\mathrm{l}} = 0.017\ \mathrm{s}$.

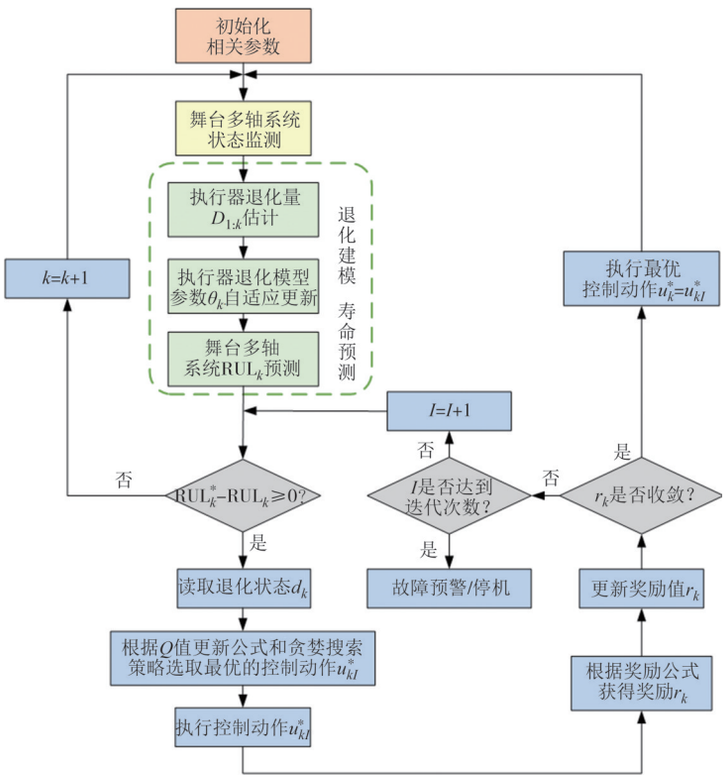


图 2 基于 Q-learning 舞台多轴同步系统 PdM 算法流程图

Fig.2 Flowchart of PdM algorithm for stage multi-axis synchronous system based on Q-learning

PID 控制层和 RL 维护层的采样时间分别取为 $T_{S1}=0.01$ s 和 $T_{S2}=0.1$ s. 采用 Ziegler-Nichols 法整定得到内环 PI 控制器参数 $k_p=0.142$、$k_I=1.38$，并取位置耦合同步补偿器为 $(0.142s+0.38)/s$，s 表示复数域的自变量. 假定在系统投运后电机才逐渐出现退化，其初始退化值 $D(0)=0$，模型参数初始值分别为 $a_0=0.985$、$b_0=3.186$ 和扩散系数 $\sigma_{B_0}=0.01$，随机噪声 $\sigma_{\eta_0}=0.003$. 根据舞台安全技术标准，设置系统失效阈值 ω_{th} 为台面位置偏差不大于 5 mm，即当台面位置偏差大于 5 mm 时，系统性能不可接受，达到寿命终点.

Q-learning 算法中超参数包括学习速率 α，折扣系数 γ，贪婪搜索速率 ε，分别设置为 $\alpha=0.5$，$\gamma=0.9$，$\varepsilon=0.2$，奖励函数中的半正定矩阵 $\boldsymbol{S}$、$\boldsymbol{R}$、$\boldsymbol{P}$ 初始值分别设置为 0.1、10、100，迭代次数 $I=2\,000$. 对比实验 DMC 的控制参数与文献[12]相同，具体参数包括：预测步长为 10，控制步长 $M=5$，误差约束矩阵的初值设置为 0.015，控制约束矩阵的初值设置为 0.04.

4.2 实验实施与结果分析

4.2.1 基于退化参数自适应更新的舞台多轴同步系统 RUL 预测

假设在仿真运行初始时刻电机性能开始下降，图 3 为未施加健康维护策略时舞台多轴同步系统的位置同步误差. 根据首达时间下失效定义，系统位置同步误差 ω 在 833 h 时刻超出失效阈值 $\omega_{\text{th}}=5$ mm，使用寿命未达到期望工作时限 $T_{\text{mission}}=900$ h，即系统在此之前失效.

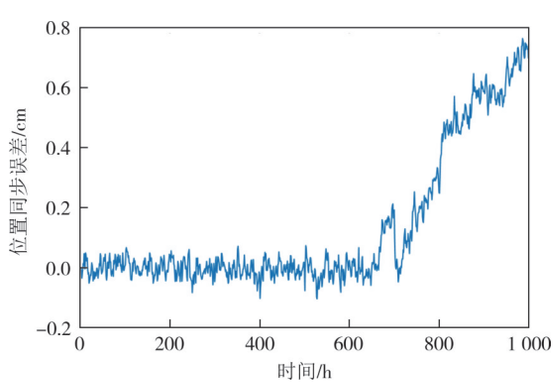


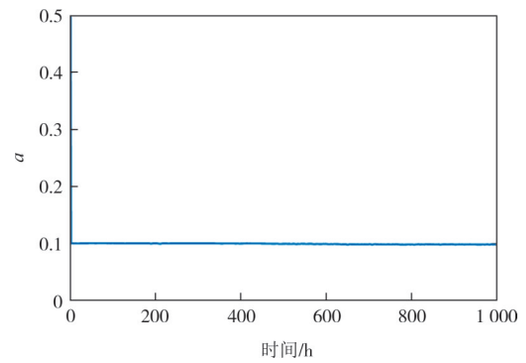
图 3 位置同步误差

Fig.3 Position synchronization error

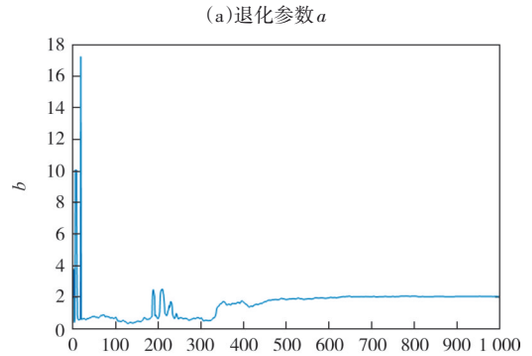
为减少退化过程的测量噪声以及维护介入、个体差异、环境等引起时变等因素影响，随着退化数据不断累积，在每个监测点通过 Kalman 滤波和 EM 算法，可实时估计退化量并自适应更新式(1)中执行器的退化参数 $\theta_k=[a_k,b_k,\sigma_{\eta_k},\sigma_{B_k}]$，如图 4 所示. 可以看出，模型参数随电机退化监测数据的积累快速

收敛,故可以此实现参数自适应实时更新，确保退化模型的精准性,进而用于系统RUL预测.

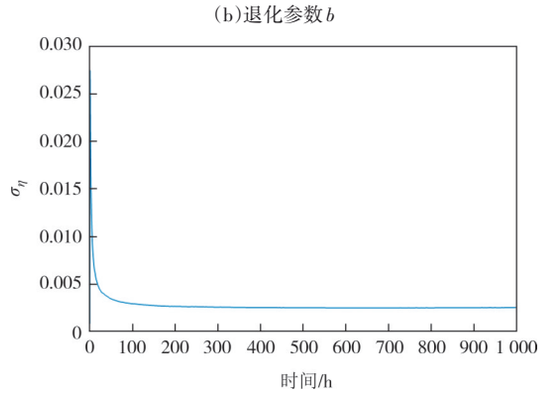
根据式(15)得到系统剩余寿命的概率密度分布PDF如图5所示.为对比本文所提方法的优越性，图中同时给出了文献[12]的结果.可以看出，随着退化监测时间的不断递增，基于本文所用模型的PDF更加尖锐，说明退化过程模型参数被自适应更新后，降低了舞台多轴系统RUL预测的不确定性，提高了寿命预测的准确性.



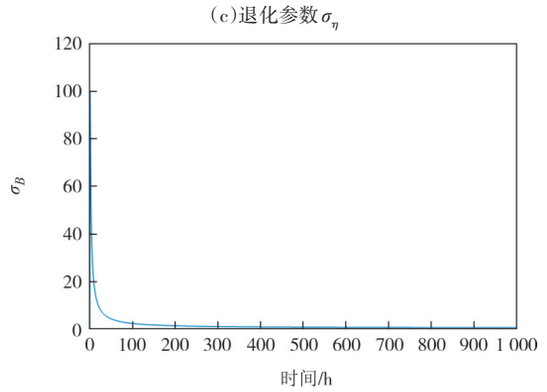
(a)退化参数a



(b)退化参数b



(c)退化参数σ_η



(d)退化参数σ_B

图4　退化参数自适应更新

Fig.4　Adaptive update of degradation parameters

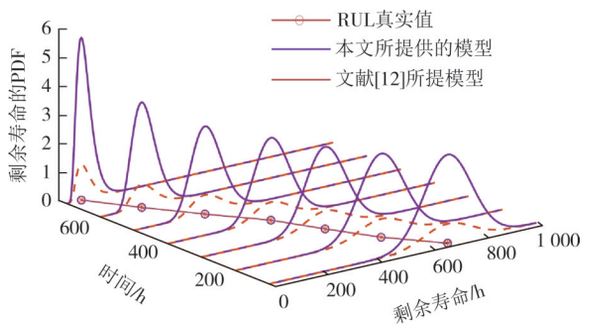


图5　系统剩余寿命预测分布

Fig.5　Distribution of system remaining useful life prediction

4.2.2　基于RL的舞台多轴同步系统自主维护策略

衡量一个RL算法的优劣通常有两个直观指标：收敛速度和累积奖励.就Q-learning扮演舞台多轴系统自主维护策略的角色而言，则需评价是否能以更快收敛速度，得到最多累积奖励的控制策略，并由实时维护获得更长的使用寿命.采用3.2节的Q-learning维护策略，当收到环境返回的奖励$R^h\left(d_k, \mathrm{RUL}_k, u_k\right)$时，紧随着进入下一状态，该过程不断反复直到算法完全收敛.图6分别给出了控制层与维护层采用恒值多采样率(T_{S1} = 0.01 s，T_{S2} = 0.1 s)和恒值单采样率(二者均取0.1 s)的维护策略收敛过程与平均奖励，共运行2 000幕,每一幕的初始值保持不变.

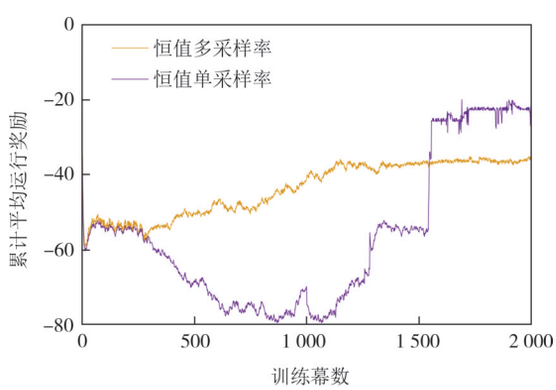


图6　2 000次迭代的累计平均运行奖励

Fig.6　Cumulative average running reward for 2 000 iterations

由图6可知，在0~300幕之间，两种采样率调控方式所获得累计平均运行奖励大致相当;在300~1 500幕之间，恒值单采样率的累计平均运行奖励总

体表现为先下行后上升,在约1 500幕之后跳变增大且高于恒值多采样率,过程表现极不稳,收敛速度也较慢;而在此区间,恒值多采样率则呈现整体逐渐上升趋势,在1 340幕左右时累计平均运行奖励收敛到约-37,之后保持在较高水平且运行较为平稳,收敛速度相对较快.表明本文针对控制层和维护层,根据其属性和需求,采用恒值多采样率的分而治之调控方式,更适合RL维护策略的实施.

在无维护和3种有维护情形下,图7和图8分别给出了控制量和电机退化量对比曲线.从图7可以看出,施加维护策略后控制量明显减小,较文献[12]的DMC维护策略,文中Q-learning的控制量在电机退化加重后减小更明显.进一步观察图8可知,由于维护策略的作用,相应的电机执行压力均得到了有效缓解,尤其伴随Q-learning维护策略对控制量的更优减弱,电机退化得到了更为显著的缓解.

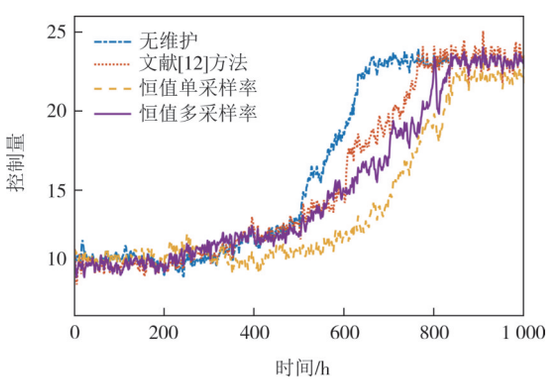


图7 有/无维护作用下控制量对比曲线

Fig.7 Comparison of control variables with/without maintenance

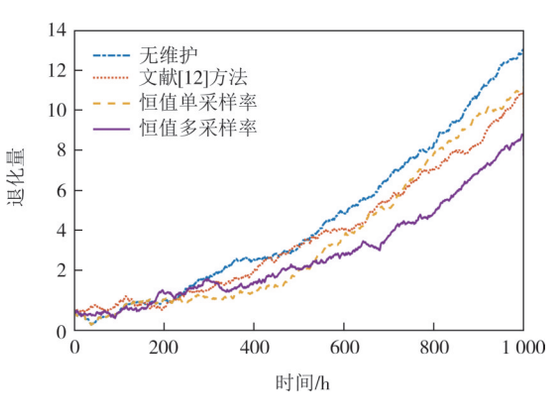


图8 有/无维护作用下退化量对比曲线

Fig.8 Comparison of degradation variables processes with/without maintenance

图9和表1分别给出了舞台多轴同步系统的位置同步误差对比曲线和系统寿命值.4种情形下,超出位置同步误差失效阈值ω_{th}=5 mm的时间分别是:无维护为833 h,文献[12]DMC维护策略为919 h,本文Q-learning维护策略的恒值单采样率为935 h、恒值多采样率为967 h.可以看出,3种维护策略均能有效延长系统寿命.进一步表明本文采用的智能维护策略通过Q-learning算法自主试错寻优,无需人工试凑调整参数,对于舞台多轴系统退化的预测维护效果更佳,采用恒值多采样率的分而治之调控方式,更适合多轴系统的工程应用.

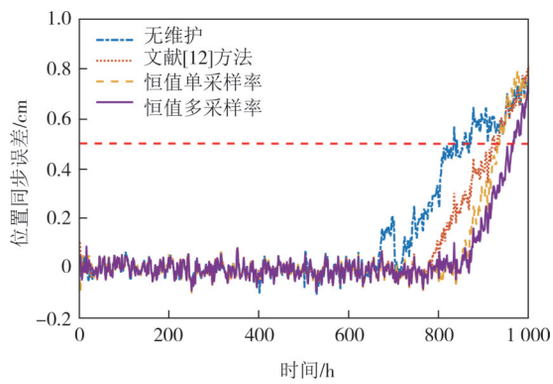


图9 有/无维护作用下位置同步误差对比曲线

Fig.9 Comparison of position synchronization errors with/without maintenance

表1 不同维护策略下系统的寿命值

Tab.1 Lifetime values of the system under different maintenance strategies

维护方法	系统寿命/h
无维护	833
文献[12]	919
恒值单采样率	935
恒值多采样率	967

5 结 论

针对单电机性能退化的舞台多轴同步控制系统,本文提出了一种基于强化学习预测维护架构的方法.通过舞台两轴同步系统仿真实验,得到以下结论:

1)结合RL和舞台健康感知信息提出的舞台多轴系统PdM架构,考虑控制层和维护层的不同需求,对其赋予不同采样率分而治之的工作方式,提高了系统维护效能;

2)考虑介入维护及多源不确定性的影响,基于Kalman滤波、EM和RTS等算法,通过对执行器时变退化模型参数的自适应更新,确保了舞台系统实时RUL预测的准确性;

3)基于系统剩余寿命偏差和执行器退化状态构建的RL算法目标函数,可通过不断试错,对维护控

制量做出最优调整,规避了MPC参数设定依赖经验的弊端,获得了使用寿命更大延长的奖励.

实际工程中,舞台多轴同步系统面临的退化情况往往更加复杂多变,考虑多个执行器同时发生退化的情况更具挑战性和实际意义,也是本文下一步拟开展的主要研究工作之一.

参考文献

[1] 克里斯蒂安·弗莱米勒.舞台机械控制系统的安全要求及轴控制技术的特点与应用[J].演艺科技,2012(2):1-4.
CHRISTIAN F. Safety requirements for stage machinery control systems and the characteristics and applications of axis control technology [J]. Entertainment Technology, 2012 (2) : 1-4. (in Chinese).

[2] 智浩,王强,金艳苓,等. 舞台机械技术、工艺、应用系列谈:舞台机械设备的多样性和发展趋势[J]. 演艺科技,2021(6):29-33.
ZHI H, WANG Q, JIN Y L, et al. Stage machinery technology, craft, application series talk: diversity and development trend of stage machinery and equipment[J]. Entertainment Technology, 2021(6):29-33.(in Chinese)

[3] 周东华,魏慕恒,司小胜. 工业过程异常检测、寿命预测与维修决策的研究进展[J]. 自动化学报,2013,39(6):711-722.
ZHOU D H, WEI M H, SI X S. A survey on anomaly detection, life prediction and maintenance decision for industrial processes[J]. Acta Automatica Sinica, 2013, 39(6): 711-722.(in Chinese)

[4] YAN J H, MENG Y, LU L, et al. Industrial big data in an industry 4.0 environment: challenges, schemes, and applications for predictive maintenance[J]. IEEE Access, 2017, 5: 23484-23491.

[5] SELCUK S. Predictive maintenance, its implementation and latest trends[J]. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 2017, 231(9): 1670-1679.

[6] 陆宁云,陈闯,姜斌,等. 复杂系统维护策略最新研究进展:从视情维护到预测性维护[J]. 自动化学报,2021,47(1):1-17.
LU N Y, CHEN C, JIANG B, et al. Latest progress on maintenance strategy of complex system: from condition-based maintenance to predictive maintenance [J]. Acta Automatica Sinica, 2021, 47(1): 1-17.(in Chinese)

[7] SALAZAR J C, WEBER P, NEJJARI F, et al. System reliability aware Model Predictive Control framework [J]. Reliability Engineering & System Safety, 2017, 167: 663-672.

[8] BOUGACHA O, VARNIER C, ZERHOUNI N. Review of post-prognostics decision-making in prognostics and health management[J]. International Journal of Prognostics and Health Management, 2021, 11(2): 1-31.

[9] LANGERON Y, GRALL A, BARROS A. Actuator health prognosis for designing LQR control in feedback systems [J]. Chemical Engineering Transactions (CET Journal), 2013, 33: 979-984.

[10] SANCHEZ H, ESCOBET T, PUIG V, et al. Health-aware model predictive control of wind turbines using fatigue prognosis [J]. IFAC-PapersOnLine, 2015, 48(21): 1363-1368.

[11] LANGERON Y, GRALL A, BARROS A. A modeling framework for deteriorating control system and predictive maintenance of actuators [J]. Reliability Engineering & System Safety, 2015, 140: 22-36.

[12] 李炜,颜伟俊,毛海杰. 多轴同步控制系统的寿命预测和延寿方法[J]. 控制与决策,2023,38(9): 2587-2596.
LI W, YAN W J, MAO H J. Research on life prediction and extension method of multi-axis synchronous control system [J]. Control and Decision, 2023, 38(9): 2587-2596.(in Chinese)

[13] 申富媛,李炜,蒋栋年. 四旋翼无人机寿命预测和自主维护方法[J]. 吉林大学学报(工学版),2023,53(3):841-852.
SHEN F Y, LI W, JIANG D N. Life prediction and self-maintenance method of quad-rotor unmanned aerial vehicle [J]. Journal of Jilin University (Engineering and Technology Edition), 2023, 53(3): 841-852.(in Chinese)

[14] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[J]. IEEE Transactions on Neural Networks, 1998, 9(5): 1054.

[15] 王龙, 黄锋. 多智能体博弈、学习与控制[J]. 自动化学报, 2023,49(3): 580-613.
WANG L, HUANG F. An interdisciplinary survey of multi-agent games, learning, and control[J]. Acta Automatica Sinica, 2023, 49(3): 580-613.(in Chinese)

[16] JHA M S, WEBER P, THEILLIOL D, et al. A reinforcement learning approach to health aware control strategy[C]//2019 27th Mediterranean Conference on Control and Automation (MED). July 1-4, 2019, Akko, Israel: IEEE, 2019: 171-176.

[17] 李天梅,司小胜,刘翔,等. 大数据下数模联动的随机退化设备剩余寿命预测技术[J]. 自动化学报,2022,48(9):2119-2141.
LI T M, SI X S, LIU X, et al. Data-model interactive remaining useful life prediction technologies for stochastic degrading devices with big data[J]. Acta Automatica Sinica, 2022, 48(9): 2119-2141.(in Chinese)

[18] KHAN S, YAIRI T. A review on the application of deep learning in system health management[J]. Mechanical Systems and Signal Processing, 2018, 107: 241-265.

[19] SI X S, REN Z Q, HU X X, et al. A novel degradation modeling and prognostic framework for closed-loop systems with degrading actuator[J]. IEEE Transactions on Industrial Electronics, 2020, 67(11): 9635-9647.

[20] 郑建飞,胡昌华,司小胜,等. 考虑不完全维护影响的随机退化设备剩余寿命预测[J]. 电子学报,2017,45(7):1740-1749.
ZHENG J F, HU C H, SI X S, et al. Remaining useful life prognostic for the stochastic degradation device subject to imperfect maintenance[J]. Acta Electronica Sinica, 2017, 45(7): 1740-1749.(in Chinese)

[21] 叶宇豪,彭飞,黄允凯. 多电机同步运动控制技术综述[J].电工技术学报,2021,36(14):2922-2935.
YE Y H, PENG F, HUANG Y K. Overview of multi-motor synchronous motion control technology[J]. Transactions of China Electrotechnical Society, 2021, 36(14): 2922-2935.(in Chinese)

[22] 路承功,魏智强,乔宏霞,等. 基于Wiener随机过程地下腐蚀环境中钢筋混凝土耐久性寿命预测[J]. 湖南大学学报(自然科学版),2021,48(7):119-128.
LU C G, WEI Z Q, QIAO H X, et al. Prediction of durability life of reinforced concrete under underground corrosion environment based on Wiener random process[J]. Journal of Hunan University (Natural Sciences), 2021, 48(7): 119-128.(in Chinese)

[23] 司小胜,胡昌华. 数据驱动的设备剩余寿命预测理论及应用[M]. 北京:国防工业出版社,2016:60-72.
SI X S, HU C H. Data-driven remaining useful life prediction theory and applications for equipment [M]. Beijing: National Defense Industry Press, 2016: 60-72.(in Chinese)