Data-Based Feedback Relearning Algorithm for Robust Control of SGCMG Gimbal Servo System with Multi-source Disturbance

ZHANG Yong[1], *MU Chaoxu*[1*], *LU Ming*[2]

1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, P. R. China;
2. Beijing Institute of Control Engineering, Beijing 100190, P. R. China

(Received 27 February 2021; revised 15 April 2021; accepted 20 April 2021)

Abstract: Single gimbal control moment gyroscope (SGCMG) with high precision and fast response is an important attitude control system for high precision docking, rapid maneuvering navigation and guidance system in the aerospace field. In this paper, considering the influence of multi-source disturbance, a data-based feedback relearning (FR) algorithm is designed for the robust control of SGCMG gimbal servo system. Based on adaptive dynamic programming and least-square principle, the FR algorithm is used to obtain the servo control strategy by collecting the online operation data of SGCMG system. This is a model-free learning strategy in which no prior knowledge of the SGCMG model is required. Then, combining the reinforcement learning mechanism, the servo control strategy is interacted with system dynamic of SGCMG. The adaptive evaluation and improvement of servo control strategy against the multi-source disturbance are realized. Meanwhile, a data redistribution method based on experience replay is designed to reduce data correlation to improve algorithm stability and data utilization efficiency. Finally, by comparing with other methods on the simulation model of SGCMG, the effectiveness of the proposed servo control strategy is verified.

Key words: control moment gyroscope; feedback relearning algorithm; servo control; reinforcement learning; multi-source disturbance; adaptive dynamic programming

CLC number: TN925 **Document code**: A **Article ID**: 1005-1120(2021)02-0225-12

0 Introduction

In the field of aerospace, the control moment gyroscope (CMG) gimbal servo system is often used as an actuator for attitude control of aerospace equipment. Fig.1 shows the typical structure of a single gimbal CMG (SGCMG) system. SGCMG has one rotor system which supports a constant angular momentum, one gimbal system that changes the angular momentum, and one structure base[1]. SGCMG changes the speed and rotation angle of the rotor system by controlling the permanent magnet synchronous motor (PMSM) in gimbal system, and then the rotor system is used as the actuator to output appropriate torque. Compared with the traditional direct control of motor drive system, SGCMG can stably provide a larger torque, which is based on the ability given by the law of conservation of angular momentum.

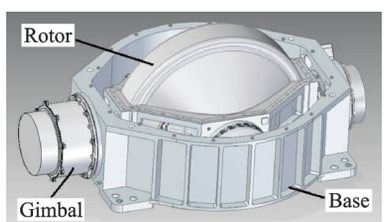

Fig.1 Typical structure of SGCMG

Under normal working condition, the output

*Corresponding author, E-mail address: cxmu@tju.edu.cn.

How to cite this article: ZHANG Yong, MU Chaoxu, LU Ming. Data-based feedback relearning algorithm for robust control of SGCMG gimbal servo system with multi-source disturbance[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2021, 38(2): 225-236.

http://dx.doi.org/10.16356/j.1005-1120.2021.02.004

torque of SGCMG system is proportional to the angular velocity of rotor system. However, in the complex space environment, the angular velocity of rotor system will be disturbed by various disturbances, which will affect the quality of the output torque. There are multi-sources for disturbance in SGCMG system, including coupled gyro torque, unbalance torque of the high speed rotor, friction torque, precision error of grating sensor, fluctuation of torque coefficient of driving motor, calculating accuracy of system circuit design[2-4]. It is worth noting that these disturbances include multi-source high-frequency, low-frequency and slope torque disturbances. Therefore, to improve the robustness and anti-interference ability of SGCMG system, some works have been conducted based on fuzzy control, sliding mode control, disturbance observer compensation, repetitive control, etc[5-9].

Many of the existing control strategies are designed based on the determined system models, to deal with high-frequency or low-frequency torque disturbances. From this perspective, when SGCMG system is faced with model uncertainty, such as torque coefficient fluctuations, the reliance on an accurate model will hinder the effectiveness of the model-based strategy and thus fail to achieve the expected control effect.

Off-policy algorithm is a kind of reinforcement learning (RL) algorithm structure that extracts model information based on system operation data, and finally obtains the control strategy without using system model[10-14]. Based on adaptive dynamic programming (ADP) method, off-policy algorithm was developed for the robust control of some linear and nonlinear systems, and the prior knowledge of the system dynamic has been relaxed[10-11]. In Ref. [12], the off-policy algorithm was extended the problem to H^{∞} control, where the ideal of integral RL (IRL) method has been applied. Considering input constraints, a two-player game problem is studied based on the off-policy algorithm in Ref. [13]. Therefore, developed from off-policy algorithm, this paper designs a feedback relearning (FR) algorithm to obtain the servo control strategy without relying on the SGCMG system model.

Considering the variability and complexity of multi-source disturbance in SGCMG system, the designed servo control strategy should have certain adaptability. In this regard, the on-policy algorithm can solve this online learning problem to improve the algorithm adaptability[15-22]. In on-policy algorithm, the obtained control strategy is rewarded or punished by designing an incentive mechanism, and then the new strategy is used to interact with the system. Continuously strengthen the control strategy to optimize the objective function, thus realizing online update and adaptive control. Therefore, the designed FR algorithm combines the idea of on-policy algorithm to realize online update of servo control strategy.

In the off-policy algorithm based on least-square principle, the collected data episodes need to satisfy certain rank conditions to ensure the validity of the matrix inverse operation. However, the correlation problem between adjacent data episodes is very serious, especially in the continuous-time robust control problem. Experience replay technology can be used to achieve faster learning by reusing the collected data[23]. The application of experience replay technology not only reduces the data correlation of the current data set, but also improves data utilization efficiency[24-25]. Meanwhile, when applying the experience replay technology to actor-critic RL algorithms, the convergence properties can also be guaranteed[26]. Therefore, referring to the idea of experience replay, a data redistribution method is designed to reduce data correlation to improve algorithm stability and data utilization efficiency.

In this paper, due to the complex mechanical structure of SGCMG, the influence of gimbal installation and the flexible support, it is difficult to obtain the accurate mathematical model in practice. The speed control problem of SGCMG is a complex servo control problem, which is also a motivation for the development of data-based RL algorithm in this paper. The data-based RL algorithm is based on the collected servo data, and the control strategy of SGCMG can be realized through iterative learning. For the convenience of problem formulation and the description of multi-source disturbance, the PMSM

model is given in section 1. In practical servo control, the controlled system is the overall SGCMG system, which is a complex nonlinear system.

The main contributions are as follows: First, inspired by on-policy and off-policy algorithms, a data-based FR algorithm is designed for the robust control problem of nonlinear system, which has the adaptability for uncertain problems and high data efficiency. Second, based on the FR algorithm, the servo control strategy of SGCMG system can be obtained by collecting servo data episodes of gimbal system. The prior knowledge of the SGCMG model is not required. Third, a data redistribution method based on experience replay is designed to reduce data correlation to further improve algorithm stability and data efficiency. Considering the multi-source disturbance, the comparison experiment with PID and SMC is given to verify the effectiveness of proposed strategy.

The main organization of this paper is as follows: Section 1 investigates the background of SGCMG gimbal servo system with multi-source disturbance. In section 2, the prior knowledge and mathematical principle of FR algorithm are described in detail. Section 3 introduces the structure of FR algorithm, the application of FR algorithm in SGCMG system, and the technology of data redistribution method based on experience replay. In section 4, the comparative simulation with other methods is analyzed. Finally, section 5 contains some conclusions of this paper.

1 Problem Formulation

SGCMG consists of one rotor system, one gimbal system and one structural base. The gimbal system is used to change the angular momentum of rotor system to output the torque. However, it is difficult to accurately express the model of SGCMG system by mathematical principle. In the existing work, we usually analyze the gimbal system, which is the driving control system, to reduce the difficulty of controller design. SGCMG is derived by controlling the PMSM, which is studied on d-q axes. Further, the state space model of PMSM is defined as[27]

$$\begin{bmatrix}\dot{I}_d(t)\\ \dot{I}_q(t)\\ \dot{w}(t)\end{bmatrix}=\begin{bmatrix}-\dfrac{R}{L_d} & pw & 0\\ -pw & -\dfrac{R}{L_q} & -\dfrac{p\varphi_{\mathrm{f}}}{L_q}\\ 0 & \dfrac{1.5P\varphi_{\mathrm{f}}}{J} & -\dfrac{f}{J}\end{bmatrix}\begin{bmatrix}I_d(t)\\ I_q(t)\\ w(t)\end{bmatrix}+\begin{bmatrix}\dfrac{u_d(t)}{L_d}\\ \dfrac{u_q(t)}{L_q}\\ -\dfrac{T_l(t)}{J}\end{bmatrix} \tag{1}$$

where the physical meaning of the model parameters are as follows. Stator current of d-q axes: I_d and I_q; d-q axes voltage: u_d and u_q; stator inductances on d-q axes: L_d and L_q; gimbal rotation speed: w; Stator resistance: R; number of pole pairs: p; flux linkage: φ_{f}; viscous friction coefficient: f; Moment of inertia: J; multi-source torque disturbance: T_l. As investigated in Ref.[4], different kinds of torque disturbances are included in T_l, including high-frequency, low-frequency and slope torque disturbances.

The multi-source disturbances can be mathematically expressed as

$$\begin{aligned}T_l(t)=&\ T_{\mathrm{G}}(\theta,w,w_h,w_{\mathrm{s}})+T_g(\theta)+T_{\mathrm{f}}(\theta,w)+\\ &\ T_{\mathrm{d}}(\theta,w_h)+T_{\mathrm{m}}(\theta)\end{aligned} \tag{2}$$

where T_{G} represents the gyroscopic effect on gyro torque; T_g the disturbance torque caused by static unbalance which will disappear when SGCMG works in aerospace; T_{f} the low-frequency torque disturbance caused by nonlinear friction of gimbal transmission parts such as bearing and conducting ring; T_{d} is related to the rotor unbalance vibration with high-frequency torque disturbance; T_{m} the high-frequency torque disturbance related to motor torque fluctuation; w_{s} the satellite speed; w_h the rotor speed; θ the gimbal angle position. The detail analysis of multi-source disturbances can be referred to Ref.[4], which will not be repeated here. It should be noted that this paper mainly focuses on the multi-source disturbances that act on the servo torque in the SGCMG system.

Remark 1 From the above description, we know that multi-source disturbances exist in SGCMG gimbal servo system, and the influence of these disturbances on high-precision servo control cannot be ignored[2-3]. However, it is still a technical bottleneck in this field to accurately describe the impact of these disturbances on the system model, which affects the construction of complete SGCMG gimbal servo system in mathematical form. Due to the differences in installation or mechanical parts, even two devices of the same type will have different model parameters. In some scenarios, these negative effects may lead to the performance degradation of model-based strategies.

Therefore, considering the difficulty of SGCMG system modeling and the influence of multi-source disturbance, a data-based FR algorithm is designed to circumvent the difficulty of accurate modeling of SGCMG. In section 2, the prior knowledge and mathematical principle of FR algorithm will be described in detail.

2 Data-Based Feedback Relearning Algorithm

2.1 Prior knowledge: On-policy and off-policy algorithm

In RL methods, on-policy and off-policy algorithms are two common algorithm structures. The core of both algorithms includes policy evaluation and policy improvement. The control strategy is evaluated based on the target indicators, and then the current strategy is improved to optimize the target function. Through continuous interaction and update of the control strategy and the system dynamics, the interactive improvement of overall strategy will eventually be achieved. Both on-policy and off-policy algorithm structures can complete the evaluation and improvement of algorithms based on the collected system data[11].

The off-policy algorithm has better data utilization efficiency and convergence ability. At first, the off-policy algorithm collects system operation data of finite dimensions and processes it into data episodes. Then, the evaluation and improvement of the control strategy can be completed through iterative learning. The collected data episodes are iterated under off-line conditions, so the off-policy algorithm does bring much burden to the storage system. At the same time, finite data episodes of off-line iteration based on least-square principle which makes the algorithm converge better and the iteration steps will be relatively small[10-13]. However, due to the characteristics of off-line iteration, the collected finite data is dynamically generated by original system, and the obtained strategy is in line with the original dynamic. Therefore, when the off-policy algorithm is used to deal with system uncertainties, the control performance may decrease.

Fortunately, the on-policy algorithm has better adaptive capabilities to uncertain systems. Based on the collected data, the control strategies can be obtained through policy evaluation and policy improvement. Different from the off-policy, the control strategy based on the on-policy algorithm is applied to the system dynamic in real time and new data will be generated for the next iteration. As a result, the collected new data episodes will contain changes in dynamic information, thereby achieving the dynamic improvement of control strategy[11,15-17]. Since the algorithm needs to constantly interact with system, the collected data are used only once in each iteration. Therefore, the on-policy algorithm performs worse than the off-policy algorithm in term of data efficiency and convergence speed.

Remark 2 On-policy and off-policy algorithms have their advantages and restrictions. On-policy algorithm has advantages in adaptability, and off-policy algorithm has advantages in convergence and data utilization. However, on-policy has low data utilization, and off-policy has insufficient adaptability to system uncertainty. Accordingly, the characteristics of FR algorithm are as follows: For the problem of multi-source disturbances, the online optimization and adaptive update of control strategy can be realized; the data redistribution method makes full use of empirical servo data; the correlation between adjacent data is reduced, and the algorithm stability and convergence are improved.

In this paper, a new algorithm structure named

FR algorithm is proposed, which has the advantages in data efficiency, algorithm convergence, and adaptability. The specific mathematical principles will be introduced in detail in section 2.2.

2.2 Mathematical principle of feedback relearning algorithm

To facilitate the introduction of mathematical principles, the unknown uncertain SGCMG system can be expressed as follows

$$\dot{\boldsymbol{x}}(t)=\boldsymbol{F}(\boldsymbol{x}(t))+\boldsymbol{G}(\boldsymbol{x}(t))(\boldsymbol{u}(t)+\boldsymbol{D}(t)) \quad (3)$$

where $\boldsymbol{x}(t)\in\boldsymbol{R}^n$ is the state vector which corresponds to the error speed $e_w(t)$ and stator current I_q of SGCMG. Define the setting speed of SGCMG as w_0, and $e_w(t)=w_0-w(t)$, $\boldsymbol{x}(t)=[e_w(t), I_q(t)]^{\mathrm{T}}$. $\boldsymbol{u}(t)\in\boldsymbol{R}^m$ represents the servo control strategy, which is related to the q axis voltage state; $\boldsymbol{D}(t)$ the multi-source disturbance of SGCMG system; $\boldsymbol{F}(x(t))$ the unknown system dynamic of SGCMG with $\boldsymbol{F}(0)=0$; and $\boldsymbol{G}(\boldsymbol{x}(t))$ the unknown control matrix.

Based on the nominal system $\dot{\boldsymbol{x}}(t)=\boldsymbol{F}(\boldsymbol{x}(t))+\boldsymbol{G}(\boldsymbol{x}(t))\boldsymbol{u}(t)$, the cost function can be defined as

$$\boldsymbol{V}(\boldsymbol{x}(t))=\int_t^{\infty}\boldsymbol{U}(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau))\mathrm{d}\tau \quad (4)$$

where $\boldsymbol{U}(\boldsymbol{x}(t),\boldsymbol{u}(t))=\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{N}\boldsymbol{x}(t)+\boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{M}\boldsymbol{u}(t)$ is the utility function with $\boldsymbol{U}(\boldsymbol{x}(0))=0$; $\boldsymbol{N}$ and $\boldsymbol{M}$ are the positive define symmetric matrices with proper dimensions n and m. The optimal cost function can be expressed as

$$\boldsymbol{V}^*(\boldsymbol{x}(t))=\min_{\boldsymbol{u}(t)\in\boldsymbol{\Omega}_u}\int_t^{\infty}\boldsymbol{U}(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau))\mathrm{d}\tau \quad (5)$$

where $\boldsymbol{\Omega}_u$ is the set of admissible control for system (3). Further, the Hamiltonian function is obtained

$$\boldsymbol{H}(\boldsymbol{x}(t),\boldsymbol{u}(t),\nabla\boldsymbol{V}(\boldsymbol{x}(t)))=\boldsymbol{U}(\boldsymbol{x}(t),\boldsymbol{u}(t))+\nabla\boldsymbol{V}(\boldsymbol{x}(t))^{\mathrm{T}}\dot{\boldsymbol{x}}(t) \quad (6)$$

where $\nabla\boldsymbol{V}(\boldsymbol{x}(t))=\partial\boldsymbol{V}(\boldsymbol{x}(t))/\partial\boldsymbol{x}(t)$ with $\boldsymbol{V}(0)=0$. Based on Bellman optimality principle, the following Hamiltonian-Jacobi-Bellman (HJB) equation can be defined as

$$0=\min_{\boldsymbol{u}(t)\in\boldsymbol{\Omega}_u}\boldsymbol{H}(\boldsymbol{x}(t),\boldsymbol{u}(t),\nabla\boldsymbol{V}(\boldsymbol{x}(t))) \quad (7)$$

where $\boldsymbol{u}^*(t)\in\boldsymbol{\Omega}_u$ represents the optimal solution of HJB equation. The optimal servo control strategy is formulated as

$$\boldsymbol{u}^*(t)=-\frac{1}{2}\boldsymbol{M}^{-1}\boldsymbol{g}(\boldsymbol{x}(t))^{\mathrm{T}}\nabla\boldsymbol{V}^*(\boldsymbol{x}(t)) \quad (8)$$

Then, substituting Eq.(8) into Eq.(7), the HJB equation will be changed as

$$0=\nabla\boldsymbol{V}(\boldsymbol{x}(t))^{\mathrm{T}}\boldsymbol{F}(\boldsymbol{x}(t))+\boldsymbol{x}(t)^{\mathrm{T}}\boldsymbol{N}\boldsymbol{x}(t)+\boldsymbol{u}(t)^{\mathrm{T}}\boldsymbol{M}\boldsymbol{u}(t)-\frac{1}{4}\nabla\boldsymbol{V}(\boldsymbol{x}(t))^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{x}(t))\boldsymbol{M}^{-1}\boldsymbol{G}(\boldsymbol{x}(t))^{\mathrm{T}}\nabla\boldsymbol{J}(\boldsymbol{x}(t)) \quad (9)$$

However, Eq.(9) is a partial differential equation, and its analytical solution is generally difficult to directly solve. Based on policy iteration (PI) algorithm, ADP was proposed to solve Eq.(9) and finally obtain an approximate solution of u^*[28]. Initialization: $\boldsymbol{V}^0(\boldsymbol{x}(0))=0$, iteration steps $i=0$, initial admissible control $\boldsymbol{u}_1(t)$.

Policy evaluation: Substitute $\boldsymbol{V}^i(\boldsymbol{x}(t))$ into Eq.(10) to get the solution of $\nabla\boldsymbol{V}^{i+1}(\boldsymbol{x}(t))$ by

$$0=\nabla\boldsymbol{V}^{i+1}(\boldsymbol{x}(t))^{\mathrm{T}}(\boldsymbol{F}(\boldsymbol{x}(t))+\boldsymbol{G}(\boldsymbol{x}(t))\boldsymbol{u}^i(t))+\boldsymbol{U}(\boldsymbol{x}(t),\boldsymbol{u}^i(t)) \quad (10)$$

Policy improvement: Update the control strategy

$$\boldsymbol{u}^{i+1}(t)=-\frac{1}{2}\boldsymbol{M}^{-1}\boldsymbol{G}^{\mathrm{T}}(\boldsymbol{x}(t))\nabla\boldsymbol{V}^{i+1}(\boldsymbol{x}(t)) \quad (11)$$

Repeat these two steps until the algorithm meets the accuracy requirements, then the corresponding servo control strategy can be obtained. During the above iteration, the model information is still needed[29]. Further, the model dynamics $\boldsymbol{F}$ and $\boldsymbol{G}$ of SGCMG system can be relaxed based on integral reinforcement learning (IRL) method[13, 16, 18].

In the iteration process, the time derivation of cost function $\boldsymbol{V}^{i+1}(\boldsymbol{x})$ can be formulated as $\mathrm{d}\boldsymbol{V}^{i+1}/\mathrm{d}t=\nabla\boldsymbol{V}^{i+1}(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{F}(\boldsymbol{x})+\boldsymbol{G}(\boldsymbol{x})(\boldsymbol{u}_1(t)+\boldsymbol{D}(t)))$, and $\boldsymbol{u}_1(t)$ is the admissible control. Under the influence of multi-source disturbances $\boldsymbol{D}(t)$, the system will not diverge during the first data collection stage. Then, define $\boldsymbol{u}^0(t)=\boldsymbol{u}_1(t)+\boldsymbol{D}(t)$.

Based on Eqs.(10, 11), we can obtain

$$\dot{\boldsymbol{V}}^{i+1}(\boldsymbol{x})=\nabla\boldsymbol{V}^{i+1}(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{F}(\boldsymbol{x})+\boldsymbol{G}(\boldsymbol{x})(\boldsymbol{u}_1(t)+\boldsymbol{D}(t))=-2\boldsymbol{u}^{i+1}(t)\boldsymbol{M}(\boldsymbol{u}^0(t)-\boldsymbol{u}^i(t))-\boldsymbol{U}(\boldsymbol{x}(t),\boldsymbol{u}^i(t)) \quad (12)$$

Integral Eq.(12) on the time interval $[t, t+\Delta t]$

$$V^{i+1}(\boldsymbol{x}(t))-V^{i+1}(\boldsymbol{x}(t+\Delta t))=\int_t^{(t+\Delta t)}2\boldsymbol{u}^{i+1}(\tau)\boldsymbol{M}(\boldsymbol{u}^0(\tau)-\boldsymbol{u}^i(\tau))\mathrm{d}\tau+\int_t^{t+\Delta t}\boldsymbol{U}(x(\tau),\boldsymbol{u}^i(\tau))\mathrm{d}\tau \tag{13}$$

Therefore, PI algorithm based on Eqs.(10, 11) has been replaced by Eq.(13), which is mathematically equivalent as the Newton's method[17]. Based on the collected data episodes, the cost function $V^{i+1}(\boldsymbol{x})$ and the control strategy $\boldsymbol{u}^{i+1}(t)$ will be solved without using system dynamics of SGCMG.

2.3 Neural network implementation based on least-square principle

In FR algorithm, the actor-critic structure neural network is used to approximate the cost function and servo control strategy of SGCMG system, and least-square principle is used in the iteration of collected data episodes. The critic network and action network can be expressed as

$$\begin{cases}V^{i,j+1}(\boldsymbol{x})=(\boldsymbol{\nu}_{\mathrm{c}}^{i,j+1})^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{c}}(\boldsymbol{x})+\varepsilon_{\mathrm{c}}^{i,j+1}\\ \boldsymbol{u}^{i,j+1}(t)=(\boldsymbol{\nu}_{\mathrm{a}}^{i,j+1})^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{a}}(\boldsymbol{x})+\varepsilon_{\mathrm{a}}^{i,j+1}\end{cases} \tag{14}$$

where $\boldsymbol{\lambda}_{\mathrm{c}}\in\boldsymbol{R}^{l_{\mathrm{c}}}$ and $\boldsymbol{\lambda}_{\mathrm{a}}\in\boldsymbol{R}^{l_{\mathrm{a}}}$ represent the activation functions; $\boldsymbol{\nu}_{\mathrm{c}}\in\boldsymbol{R}^{l_{\mathrm{c}}\times m_{\mathrm{c}}}$ and $\boldsymbol{\nu}_{\mathrm{a}}\in\boldsymbol{R}^{l_{\mathrm{a}}\times m_{\mathrm{a}}}$ the ideal weights of critic and action networks; l_{c} and l_{a} the neuron number in hidden layer. The reconstruction errors ε_{c} and ε_{a} can be omitted as the number of iteration steps large enough[30-31]. Therefore, define the estimated form of Eq.(14)

$$\begin{cases}\hat{V}^{i,j+1}(\boldsymbol{x})=(\hat{\boldsymbol{\nu}}_{\mathrm{c}}^{i,j+1})^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{c}}(\boldsymbol{x})\\ \hat{\boldsymbol{u}}^{i,j+1}(t)=(\hat{\boldsymbol{\nu}}_{\mathrm{a}}^{i,j+1})^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{a}}(\boldsymbol{x})\end{cases} \tag{15}$$

where i and j represent the iterative steps in the outer loop and the inner loop, respectively. For example, $u^{i,j+1}(x)$ represents the $(j+1)$th iteration solution of the inner loop in the ith outer loop. The structure of algorithm iteration will be introduced in section 3. Define a large time sequence $\{t_k, k\in(0,\cdots,q)\}$, and q is the dimension requirement in data collection which satisfies $q\geqslant l_{\mathrm{c}}+l_{\mathrm{a}}m_{\mathrm{a}}$ to meet the full rank condition in the matrix inverse operation[13].

Based on Eqs.(13, 15), the residual error $\boldsymbol{\varsigma}$ is formulated as

$$\begin{aligned}\boldsymbol{\varsigma}_k^{i,j+1}(t)=&\hat{\boldsymbol{V}}^{i,j+1}(\boldsymbol{x}_k)-\hat{\boldsymbol{V}}^{i,j+1}(\boldsymbol{x}_{k+1})-\\ &2\int_{t_k}^{t_{k+1}}\{\hat{\boldsymbol{u}}^{i,j+1}\boldsymbol{M}(\boldsymbol{u}^0-\boldsymbol{u}^{i,j})\}\mathrm{d}\tau-\\ &\int_{t_k}^{t_{k+1}}\boldsymbol{U}(\boldsymbol{x},\boldsymbol{u}^{i,j})\mathrm{d}\tau=\\ &(\boldsymbol{\lambda}_{\mathrm{c}}(\boldsymbol{x}_k)-\boldsymbol{\lambda}_{\mathrm{c}}(\boldsymbol{x}_{k+1}))^{\mathrm{T}}\hat{\boldsymbol{w}}_{\mathrm{c}}^{i,j+1}-\\ &2\int_{t_k}^{t_{k+1}}\{\boldsymbol{\lambda}_{\mathrm{a}}^{\mathrm{T}}(\boldsymbol{x})\hat{\boldsymbol{w}}_{\mathrm{a}}^{i,j+1}\boldsymbol{M}(\boldsymbol{u}^0-(\hat{\boldsymbol{w}}_{\mathrm{a}}^{i,j})^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{a}}(\boldsymbol{x}))\}\mathrm{d}\tau-\\ &\int_{t_k}^{t_{k+1}}\{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}+\boldsymbol{\lambda}_{\mathrm{a}}^{\mathrm{T}}(\boldsymbol{x})\hat{\boldsymbol{\nu}}_{\mathrm{a}}^{i,j}\boldsymbol{M}(\hat{\boldsymbol{\nu}}_{\mathrm{a}}^{i,j})^{\mathrm{T}}\boldsymbol{\lambda}_{\mathrm{a}}(\boldsymbol{x})\}\mathrm{d}\tau\end{aligned} \tag{16}$$

where $\boldsymbol{\varsigma}$ is introduced in the process of neural network approximation. The purpose of iteration is to get the optimal weights of neural networks, so that the residual error will converge to the minimum value. Then, Eq.(16) can be expressed as

$$\boldsymbol{\varsigma}_k^{i,j+1}(t)=\boldsymbol{\Xi}_k^{i,j}\mathrm{vec}(\boldsymbol{\varpi}^{i,j+1})-\boldsymbol{\Theta}^{i,j} \tag{17}$$

where

$$\begin{aligned}\boldsymbol{\Xi}_k^{i,j}=&[(\boldsymbol{\lambda}_{\mathrm{c}}(\boldsymbol{x}_k)-\boldsymbol{\lambda}_{\mathrm{c}}(\boldsymbol{x}_{k+1}))^{\mathrm{T}}\cdot\\ &2\int_{t_k}^{t_{k+1}}(\boldsymbol{\lambda}_{\mathrm{a}}(\boldsymbol{x})^{\mathrm{T}}\otimes\boldsymbol{\lambda}_{\mathrm{a}}(\boldsymbol{x})^{\mathrm{T}})(\hat{\boldsymbol{\nu}}_{\mathrm{a}}^{i,j}\boldsymbol{M}\otimes\boldsymbol{I}_{\lambda_{\mathrm{a}}})\mathrm{d}\tau-\\ &2\int_{t_k}^{t_{k+1}}((\boldsymbol{u}^0)^{\mathrm{T}}\otimes\boldsymbol{\lambda}_{\mathrm{a}}(\boldsymbol{x})^{\mathrm{T}})(\boldsymbol{M}\otimes\boldsymbol{I}_{\lambda_{\mathrm{a}}})\mathrm{d}\tau]\end{aligned} \tag{18}$$

and

$$\begin{aligned}\boldsymbol{\Theta}_k^{i,j}=&\int_{t_k}^{t_{k+1}}\{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}+(\boldsymbol{\lambda}_{\mathrm{a}}(\boldsymbol{x})^{\mathrm{T}}\otimes\boldsymbol{\lambda}_{\mathrm{a}}(\boldsymbol{x})^{\mathrm{T}})\times\\ &\mathrm{vec}(\hat{\boldsymbol{\nu}}_{\mathrm{a}}^{i,j}\boldsymbol{M}(\hat{\boldsymbol{\nu}}_{\mathrm{a}}^{i,j})^{\mathrm{T}})\}\end{aligned} \tag{19}$$

where $\boldsymbol{I}_{\lambda_{\mathrm{a}}}$ is the identity matrix with appropriate dimension; $\mathrm{vec}(\boldsymbol{A})$ the column vector representation of matrix $\boldsymbol{A}$, where all the column vectors are placed in one column. $\boldsymbol{\varpi}^{i,j+1}=[\mathrm{vec}(\hat{\boldsymbol{\nu}}_{\mathrm{c}}^{i,j+1})^{\mathrm{T}},\mathrm{vec}(\hat{\boldsymbol{\nu}}_{\mathrm{a}}^{i,j+1})^{\mathrm{T}}]^{\mathrm{T}}$; $\otimes$ the Kronecker Product operation.

In the ith iteration, the collected data set can be defined as

$$\boldsymbol{\Xi}^{i,j}=[(\boldsymbol{\Xi}_0^{i,j})^{\mathrm{T}},(\boldsymbol{\Xi}_1^{i,j})^{\mathrm{T}},\cdots,(\boldsymbol{\Xi}_q^{i,j})^{\mathrm{T}}]^{\mathrm{T}} \tag{20}$$

and

$$\boldsymbol{\Theta}^{i,j}=[\boldsymbol{\Theta}_0^{i,j},\boldsymbol{\Theta}_1^{i,j},\cdots,\boldsymbol{\Theta}_q^{i,j}]^{\mathrm{T}} \tag{21}$$

Based on the least-square principle, the weight parameters can be calculated by

$$\boldsymbol{\varpi}^{i,j+1}=[(\boldsymbol{\Xi}^{i,j})^{\mathrm{T}}\boldsymbol{\Xi}^{i,j}]^{-1}(\boldsymbol{\Xi}^{i,j})^{\mathrm{T}}\boldsymbol{\Theta}^{i,j} \tag{22}$$

Therefore, based on the neural network implementation and least-square principle, system model is not needed in the proposed servo control strategy, which circumvent the difficulty of SGCMG modeling.

3 Algorithm Structure and Data Redistribution

To solve the problem of multi-source disturbance, the servo control strategy obtained in FR algorithm interacts with SGCMG system and realizes the adaptive adjustment based on RL. The basic structure of FR algorithm is shown in Fig.2, including the outer loop and the inner loop iterations.

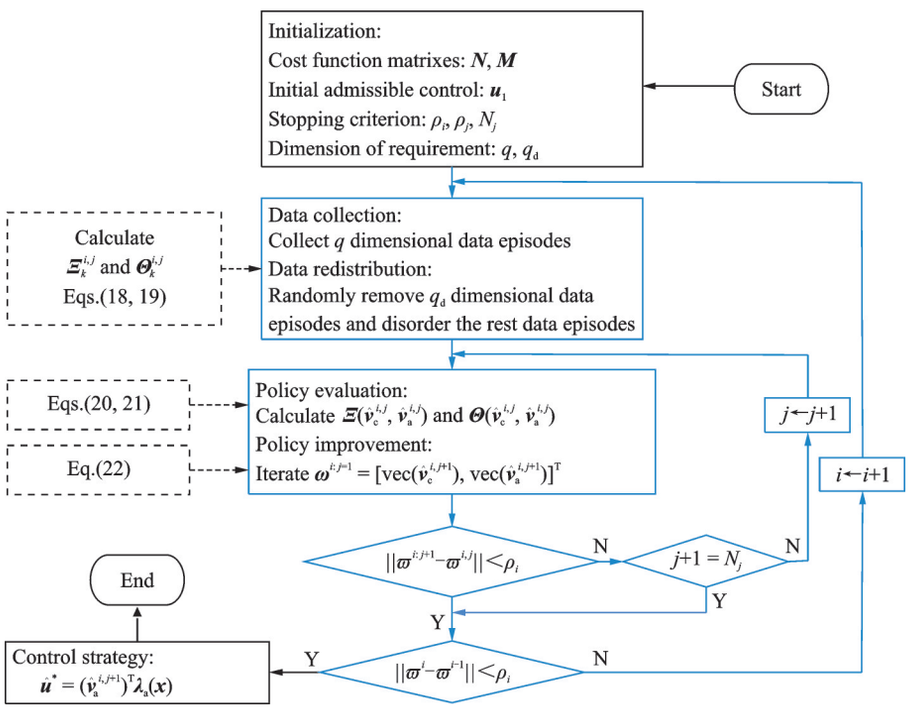

Fig.2 Typical structure of SGCMG

The first step is algorithm initialization, which involves the parameter initialization and system operation of SGCMG. Further, the algorithm collects the servo status of gimbal system in real time, performs calculations according to Eqs.(18, 19) and stores them in the memory pool until the algorithm dimension requirement q is satisfied. Then, q_d dimensional data will be randomly deleted by using the data redistribution method. Based on the least-square principle, the inner loop iteration is performed based on Eqs.(20—22). Until the calculation accuracy ρ_j is satisfied or the maximum number of iteration N_j is reached, then the outer loop criterion is performed. When the accuracy ρ_i is satisfied, the corresponding servo control strategy can be obtained. Meanwhile, the pseudo code of FR algorithm is given in Algorithm 1.

Algorithm 1 FR Algorithm

```
Start
    Initialization:
    Data collection:
        If q is satisfied
            Collect speed error states of gimbal system;
            Calculate data episodes;
            Store data episodes in the memory pool;
        End if
    Data redistribution:
        Randomly remove qd dimensional data episodes in q;
    Policy evaluation and improvement:
        Do least-square iteration based on Eqs.(20—22);
        While ρj or Nj is satisfied
        If ρi is satisfied
            Obtain the servo control strategy;
        Else
            Return to data collection step;
        End if
End
```

It is worth noting that for the data-based RL method, the uncertain data episodes will affect the

algorithm convergence. In this paper, $\boldsymbol{D}(t)$ related to the multi-source disturbances will directly affect the accuracy of collected data episodes and the dynamic performance. In this scenario, the high correlation of collected data is an important factor, which may promote the singularity of matrix operation.

Based on experience replay technology, a data redistribution method is designed to effectively reduce the correlation of collected data, and then improve the convergence performance and data utilization of FR algorithm. In the iteration of FR algorithm, the collected data episodes will be preprocessed before each inner loop iteration. In order to reduce the correlation between data episodes, q_{d} dimensional data episodes in the data set will be randomly eliminated, and the sequence of the rest episodes will be disordered.

In the face of uncertain system, this processing will be beneficial to the convergence of data-based algorithm, so as to improve the stability of the algorithm. In the next outer loop iteration, the last data set can still be retained, and only the episodes eliminated in advance need to be supplemented to meet the iteration requirements q. This can greatly improve the efficiency of data utilization, and it is also the advantage of the proposed data redistribution method.

4 Simulation Analysis

In this paper, multi-source disturbance is considered in a simulation model of SGCMG gimbal servo system. The parameters of SGCGM are given in Table 1.

In simulation, PID and SMC methods are used to compare with the servo control strategy based on FR algorithm, which is called FR control. The PID controllers are listed in Table 2, and the SMC method can be found in Ref.[4].

Table 1 SGCMG gimbal servo system parameters

Parameter	Value
Motor stator resistance/Ω	$R=2.875$
Motor stator inductance/mH	$L_d=8.5, L_q=8.5$
Flux linkage	$\varphi_{\mathrm{f}}=0.175$
Motor pole pairs	$p=6$
Moment of inertia/(kg•m^2)	$J=1.1$
Viscous friction coefficient/(Nm•s•(°)$^{-1}$)	$f=0.1$
Position sensor resolution/bit	21

Table 2 Parameters of PID controller

Controller	Value
Position PD controller	$35+0.1s/(1+0.001s)$
Speed PI controller	$5+0.01/s$
Current PI controller	$20+0.001/s$

For the training process of FR control, the parameters is set as: $\boldsymbol{N}=2\times\boldsymbol{I}^{2\times2}$, $\boldsymbol{M}=\boldsymbol{I}$ ($\boldsymbol{I}$ is the identity matrix), $\rho_i=\rho_j=1\times10^{-6}$, $N_j=100$, $q=100$ and $q_{\mathrm{d}}=40$. The activation function of action and critic networks are $\boldsymbol{\lambda}_{\mathrm{a}}(\boldsymbol{x})=\boldsymbol{\lambda}_{\mathrm{c}}(\boldsymbol{x})=[e_w^2, e_wI_q, I_q^2, e_w, I_q]^{\mathrm{T}}$. Then, the weight training processes of two networks are shown in Figs. 3, 4. If the on-policy algorithm is used, the weight parameters after one iteration will be applied to SGCMG. However, the current parameters have not converged to the optimal solution, and cannot ensure the stability of SGCMG under multi-source disturbance.

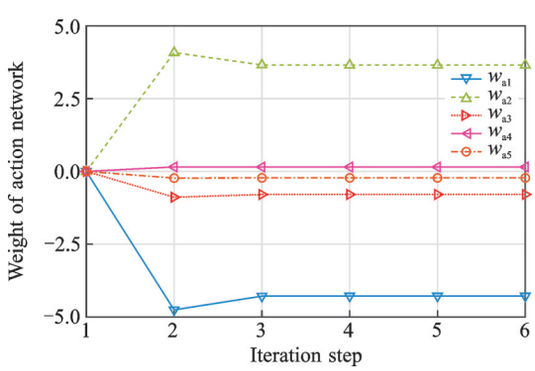

Fig.3 Weight training process of action network

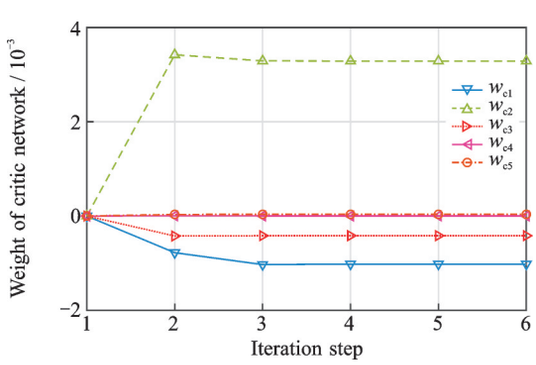

Fig.4 Weight training process of critic network

Based on the definition of state variable $\boldsymbol{x}(t)=[e_w(t), I_q(t)]^{\mathrm{T}}$, the collected data of SGCMG system are given in Fig.5. More importantly, the weights of neural networks are iterated from 0, where the selection of the initial weights in the iterative algorithm is relaxed, and it is more conducive to engineering.

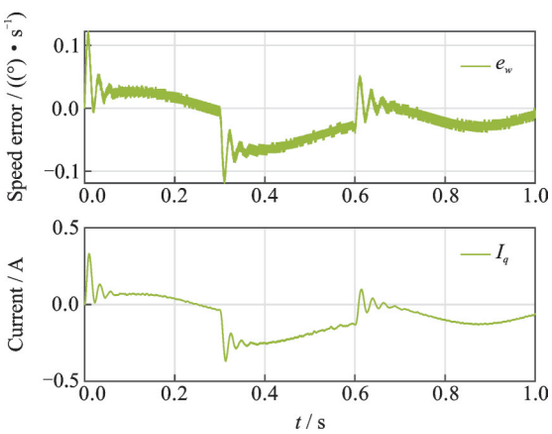

Fig.5 Collected data of SGCMG system

Fig.6 shows the training process of FR control strategy, including data collection process under admissible control (before 0.5 s), algorithm iteration (at 0.5 s), and the control process (after 0.5 s). The sampling time of SGCMG system is set as 0.005 s. Combined with the requirement of $q=100$, the data collection process lasted for 0.5 s. The algorithm iterates for a short time at 0.5 s, and then outputs the servo control strategy.

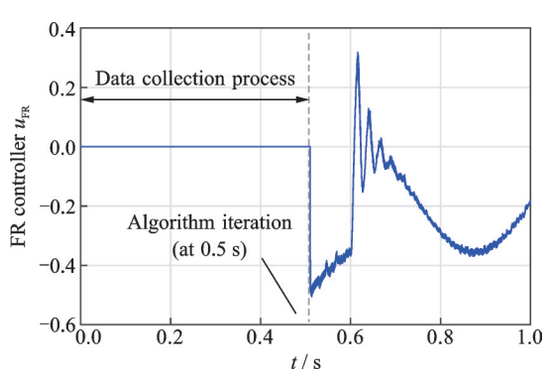

Fig.6 Training process of FR controller

The multi-source disturbance including high-frequency, low-frequency sinusoidal disturbances and slope disturbance have been shown in Fig.7, which are used to simulate precision errors introduced by position sensors, the unbalance torque of the high-speed rotor, coupled torque by satellite speed, etc.

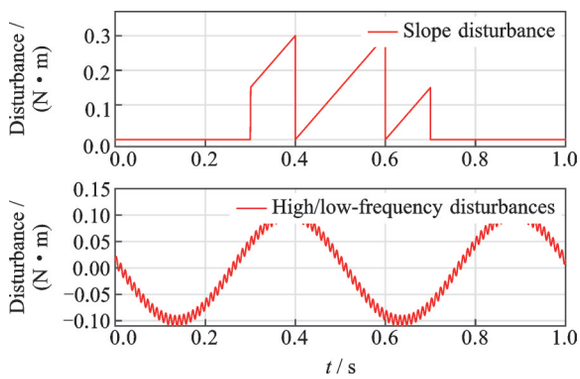

Fig.7 Multi-source disturbance

Then, Fig.8 gives the tracking control of the SGCMG system under the multi-source disturbance. The control signal of FR controller is given in Fig.9. In the simulation, the setting speed is set as $w_0=0.5\,°/\mathrm{s}$. It can be observed that SMC and PID control are greatly affected to some extent under this complex disturbance. In contrast, the FR controller shows better control performance in stability and rapidity. Based on the proposed FR algorithm, there is still a small fluctuation in the speed output. However, the robustness of SGCMG system is obviously improved, and the speed can converge to the expected value faster. At the same time, the strategy proposed in this paper is a model-free method based on data collection, which also improves the generalization of the control strategy.

The correlation of adjacent data is shown in Fig.10. Based on Eqs.(20, 21), "Data 1" repre-

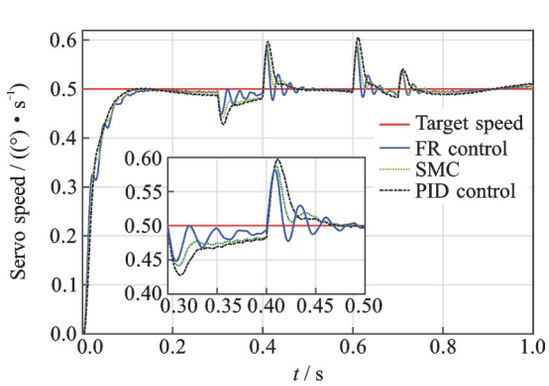

Fig.8 Servo speed control under multi-source disturbance

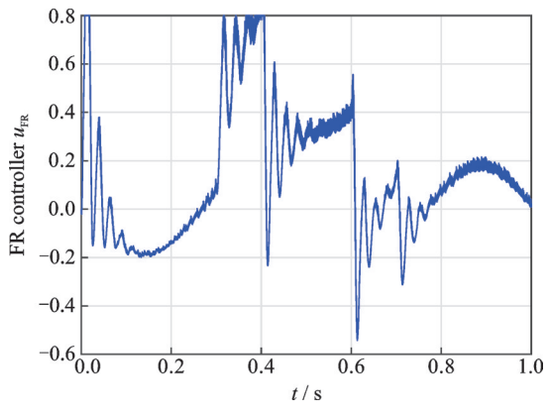

Fig.9 FR control signal under multi-source disturbance

sents the collected data set in the first iteration with $i=1$, i.e., $\boldsymbol{\Xi}^{1,1}$ and $\boldsymbol{\Theta}^{1,1}$, where the data redistribution method has not been used. Accordingly, "Data 2" is the collected data set in the second iteration with $i=2$. It indicates that the data redistribution method has been used. Fig.10 shows that the correlation of adjacent data is significantly reduced by data redistribution. For the data-based RL algorithm, high correlation of adjacent data may lead to poor convergence or even divergence of algorithm. Therefore, the data redistribution method can reduce the data correlation and improves the convergence performance of FR algorithm.

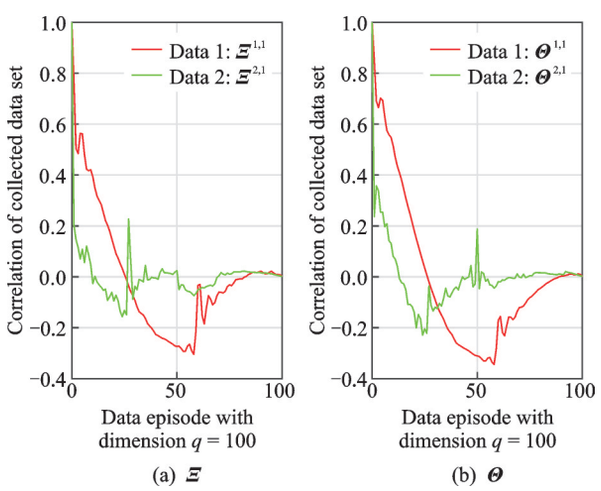

Fig.10 Correlation of collected data set

5 Conclusions

A data-based FR algorithm is proposed for the robust control of SGCMG gimbal servo system, where a data redistribution method is designed to improve the data utilization and algorithm convergence. Under the influence of multi-source disturbance, the control strategy can be obtained by using the collected data of SGCMG. This method avoids the difficulty of mathematical modeling of SGCMG and has better adaptability for uncertain problems. Through the comparative analysis on simulation platform, the proposed method can better suppress the multi-source disturbance, in terms of rapidity and stability.

References

[1] WEI D, LI G, RONG F, et al. Design of SGCMG and long life rotor bearing system technology in tiangong-1[J]. Scientia Sinica Technologica, 2014, 44(3): 261-268.

[2] LU M, LI Y, ZHANG J, et al. Ultra-low speed detection method for CMG gimbal servo systems[J]. Journal of Chinese Inertial Technology, 2012, 20(2): 234-238.

[3] SU Y, ZHENG C, MUELLER P C, et al. A simple improved velocity estimation for low-speed regions based on position measurements only[J]. IEEE Transactions on Control Systems Technology, 2006, 14(5): 937-942.

[4] LU M, WANG Y, HU Y, et al. Composite controller design for PMSM direct drive SGCMG gimbal servo system[C]//Proceedings of 2017 IEEE International Conference on Advanced Intelligent Mechatronics (AIM). Munich, Germany: IEEE, 2017: 106-112.

[5] LI J, LIU G, LI H. Fuzzy control method of gimbal servo system in magnetically suspended rotor system of CMG[J]. Micromotors, 2009, 42(5): 35-38.

[6] LU M, LI X, LI Y. Attenuation of wide margin disturbance fluctuation in SGCMG gimbal servo system[C]//Proceedings of 2012 15th International Conference on Electrical Machines and Systems (ICEMS). Sapporo, Japan: [s.n.], 2012: 1-4.

[7] LU M, ZHANG X, LI Y. Analysis and control of disturbance torque in SGCMG gimbal servo system[J]. Chinese Space Science and Technology, 2013, 1: 15-20.

[8] ZHANG J, ZHOU D, GAO Y. Gimbal control technique and gimbal rate measurement method for the control moment gyro[J]. Aerospace Control and Application, 2008, 34(2): 23-28.

[9] RAN C K. Research on a repetitive control method of brushless DC motor[J]. Micromotors, 2009, 42(7): 83-84.

[10] JIANG Y, JIANG Z P. Computational adaptive opti-

mal control for continuous-time linear systems with completely unknown dynamics[J]. Automatica, 2012, 48(10): 2699-2704.

[11] JIANG Y, JIANG Z P. Robust adaptive dynamic programming[M]. USA: John Wiley & Sons, 2017.

[12] LUO B, WU H N, HUANG T. Off-policy reinforcement learning for H_∞ control design[J]. IEEE Transactions on Cybernetics, 2015, 45(1): 65-76.

[13] ZHANG Q, ZHAO D, ZHU Y. Data-driven adaptive dynamic programming for continuous-time fully cooperative games with partially constrained inputs[J]. Neurocomputing, 2017, 238: 377-386.

[14] YANG X, HE H, LIU D. Event-triggered optimal neuro-controller design with reinforcement learning for unknown nonlinear systems[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019, 49(9): 1866-1878.

[15] VAMVOUDAKIS K G, LEWIS F L. Online solution of nonlinear two-player zero-sum games using synchronous policy iteration[J]. International Journal of Robust and Nonlinear Control, 2012, 22(13): 1460-1483.

[16] VRABIE D, LEWIS F L. Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems[J]. Neural Networks, 2009, 22(3): 237-246.

[17] LUO B, WU H N. Computationally efficient simultaneous policy update algorithm for nonlinear H_∞ state feedback control with galerkins method[J]. International Journal of Robust and Nonlinear Control, 2013, 23(9): 991-1012.

[18] VRABIE D, PASTRAVANU O, ABU-KHALAF M, et al. Adaptive optimal control for continuous-time linear systems based on policy iteration[J]. Automatica, 2009, 45(2): 477-484.

[19] HE H, NI Z, FU J. A three-network architecture for on-line learning and optimization based on adaptive dynamic programming[J]. Neurocomputing, 2012, 78(1): 3-13.

[20] ZHANG H, WEI Q, LIU D. An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games[J]. Automatica, 2011, 47(1): 207-214.

[21] MU C, ZHANG Y, GAO Z, et al. ADP-based robust tracking control for a class of nonlinear systems with unmatched uncertainties[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50(11): 4056-4067.

[22] MU C, ZHANG Y. Learning-based robust tracking control of quadrotor with time-varying and coupling uncertainties[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(1): 259-273.

[23] ADAM S, BUSONIU L, BABUSKA R. Experience replay for real-time reinforcement learning control[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2011, 42(2): 201-212.

[24] LIN L J. Self-improving reactive agents based on reinforcement learning, planning and teaching[J]. Machine learning, 1992, 8(3/4): 293-321.

[25] MODARES H, LEWIS F L, NAGHIBI-SISTANI M B. Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems[J]. Automatica, 2014, 50(1): 193-202.

[26] WAWRZYNSKI P. Real-time reinforcement learning by sequential actor-critics and experience replay[J]. Neural Networks, 2009, 22(10): 1484-1497.

[27] HEBERTT S R, JESÚS L F, CARLOS G R, et al. On the control of the permanent magnet synchronous motor: An active disturbance rejection control approach[J]. IEEE Transactions on Control Systems Technology, 2014, 22(5): 2056-2063.

[28] WERBOS P J. Approximate dynamic programming for real-time control and neural modeling, handbook of intelligent control[M]. [S.l.]: Van Nostrand Reinhold, 1992: 493-526.

[29] SUTTON R S, BARTO A G. Reinforcement learning: An introduction[M]. USA: MIT Press, 2018.

[30] FINLAYSON B A. The method of weighted residuals and variational principles[M]. New York, USA: Academic Press, 1972.

[31] MU C, WANG D, HE H. Novel iterative neural dynamic programming for data-based approximate optimal control design[J]. Automatica, 2017, 81: 240-252.

Acknowledgements This work was supported by the National Natural Science Foundation of China(No.62022061), Tianjin Natural Science Foundation(No.20JCYBJC00880), and Beijing Key Laboratory Open Fund of Long-Life Technology of Precise Rotation and Transmission Mechanisms.

Authors Dr. **ZHANG Yong** received the B.S. degree in automation from Tianjin University, Tianjin, China, in 2017, where he is currently pursuing the Ph.D. degree in control engineering with the School of Electrical and Information Engineering. His research interests include adaptive and robust control, adaptive dynamic programming, model-free con-

trol, neural networks, and related applications.

Prof. **MU Chaoxu** received the Ph.D. degree in control science and engineering from the School of Automation, Southeast University, Nanjing, China, in 2012. She was a visiting Ph.D. student with the Royal Melbourne Institute of Technology University, Melbourne, VIC, Australia, from 2010 to 2011. She was a postdoctoral fellow with the Department of Electrical, Computer and Biomedical Engineering, the University of Rhode Island, Kingston, RI, USA, from 2014 to 2016. She is currently a professor with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. She has authored more than 100 journal and conference papers, and coauthored two monographs. Her current research interests include nonlinear system control and optimization, adaptive and learning systems.

Author contributions Dr. **ZHANG Yong** designed and implemented the algorithm, compiled the model, completed the simulation analysis, and wrote the manuscript. Prof. **MU Chaoxu** provided design ideas, participated in algorithm design, and provided data analysis. Dr. **LU Ming** provided the model composition, contributed to the background and data analysis of the study. All authors commented on the manuscript draft and approved the submission.

Competing interests The authors declare no competing interests.

(Production Editor: ZHANG Bei)

基于数据的多源干扰SGCMG框架伺服系统鲁棒控制反馈再学习算法

张 勇[1], 穆朝絮[1], 鲁 明[2]

(1.天津大学电气自动化与信息工程学院, 天津 300072, 中国; 2.北京控制工程研究所, 北京 100190, 中国)

摘要: 单框架控制力矩陀螺(Single gimbal control moment gyroscope, SGCMG)具有高精度、快速响应的特点,是航天领域高精度对接、快速机动导航和制导系统的重要姿态控制系统。本文考虑多源干扰的影响,设计了一种基于数据的反馈再学习(Feedback relearning, FR)算法,用于SGCMG框架伺服系统的鲁棒控制。基于自适应动态规划和最小二乘原理,通过采集SGCMG系统的在线运行数据,采用FR算法得到伺服控制策略。这是一种无模型学习策略,无须事先了解SGCMG模型。进而,基于强化学习机制将伺服控制策略与SGCMG系统动态相互作用,可以实现伺服控制策略对多源干扰的自适应评估和改进。同时,设计了一种基于经验回放的数据重分配方法,降低了数据相关性,提高了算法稳定性和数据利用率。最后,在SGCMG仿真模型上与其他方法进行了比较,验证了所提出的伺服控制策略的有效性。

关键词: 控制力矩陀螺;反馈再学习算法;伺服控制;强化学习;多源干扰;自适应动态规划