应用数学和力学
APPLIED MATHEMATICS AND MECHANICS

基于actor-critic算法的分数阶多自主体系统最优主-从一致性控制

马丽新，刘　晨，刘　磊

Optimal Leader-Following Consensus Control of Fractional-Order Multi-Agent Systems Based on the Actor-Critic Algorithm

MA Lixin, LIU Chen, and LIU Lei

在线阅读 View online: https://doi.org/10.21656/1000-0887.420124

http://www.applmathmech.cn

基于 actor-critic 算法的分数阶多自主体系统最优主-从一致性控制*

马丽新，　刘　晨，　刘　磊

（河海大学 理学院，南京 211100）

摘要：　研究了分数阶多自主体系统的最优主-从一致性问题. 在考虑控制器周期间歇的前提下, 将分数阶微分的一阶近似逼近式、事件触发机制和强化学习中的 actor-critic 算法有机整合, 设计了基于周期间歇事件触发策略的强化学习算法结构. 最后, 通过数值仿真实验证明了该算法的可行性和有效性.

关　键　词：　分数阶多自主体系统；　actor-critic 算法；　最优主-从一致性；　事件触发；　间歇

中图分类号：　TP273; O232　　**文献标志码：**　A　　DOI：10.21656/1000-0887.420124

Optimal Leader-Following Consensus Control of Fractional-Order Multi-Agent Systems Based on the Actor-Critic Algorithm

MA Lixin，　LIU Chen，　LIU Lei

（*College of Science*, *Hohai University*, *Nanjing* 211100, *P.R.China*）

Abstract: Aimed at the optimal leader-following consensus problem of fractional-order multi-agent systems, an reinforcement learning strategy was designed based on the intermittent event trigger. With the periodic intermittent strategy as the basic mechanism, the event trigger and the actor-critic algorithm in reinforcement learning were organically integrated. According to the 1st-order approximation of the fractional differential, the reinforcement learning algorithm structure based on the periodic intermittent event trigger strategy was proposed. Finally, the feasibility and effectiveness of the algorithm was proved through numerical simulation experiments.

Key words: fractional-order multi-agent system; actor-critic algorithm; optimal leader-following consensus; event trigger; intermittence

引　　言

多自主体系统的分布式协同控制广泛存在于自然界中, 如鱼群、蜂群、鸟群等, 近年来, 在生物系统、传感器网络、无人机编队、机器人团队、水下机器人[1-4]等领域被大范围应用. 一致性是多自主体系统分布式协同

* **收稿日期：**　2021-05-07; **修订日期：**　2021-12-03

基金项目：　国家自然科学基金(面上项目)(61773152); 中央高校基本科研业务费(2019B19214)

作者简介：　马丽新(1997—), 女, 硕士生 (E-mail: 1623406486@qq.com);
刘晨(1993—), 男, 博士生 (E-mail: liuchen_hhu@163.com);
刘磊(1983—), 男, 副教授, 博士生导师 (通讯作者. E-mail: liulei_hust@163.com).

引用格式：　马丽新, 刘晨, 刘磊. 基于 actor-critic 算法的分数阶多自主体系统最优主-从一致性控制[J]. 应用数学和力学, 2022, **43**(1): 104-114.

控制的基本问题之一, 即多自主体在某种适当的协议下收敛到一个共同的状态. 2002 年, 系统与控制领域的学者 Fax 和 Murray 首次运用控制理论的观点证明, 通过对每个智能体设计仅仅依赖个体间局部信息交互的分布式控制协议, 就能驱动整个多智能体系统完成状态一致的控制目标, 并推导出一致性条件[5]. 后又有众多学者针对多自主体系统的一致性展开了研究[6-9].

由于分数阶微积分是整数阶微积分的推广, 而且近年来越来越多的研究表明: 众多实际系统运用分数阶模型才能反映出其更好的性质 (黏弹性、记忆与遗传特性等). 所以, 分数阶系统的相关研究引起国内外学者的广泛关注. 随着分数阶系统逐渐被重视, 节点带有分数阶动力学网络系统的一致性逐渐成为当下的热点问题之一,可参考文献 [10-12].

随着网络技术的发展, 考虑到通信带宽、资源利用率等问题, Astrom 等在文献 [13] 中提出事件触发控制技术以减少信息传递与调整控制器的次数. 2009 年, Dimarogonas 等[14] 将事件触发机制引入到了多智能体系统. 2014 年, Xu 等[15] 提出事件触发控制策略来研究分数阶多自主体系统的一致性问题. 2017 年, Wang 等[16] 研究了基于指数型事件触发策略下的非线性分数阶多自主体系统的主-从一致性. 此外, 间歇控制策略因具有减少控制器持续运作时间的功能, 对于解决实际工程上控制器设备限制等问题上有一定优势, 近年来被越来越多的学者应用到分数阶多自主体系统的一致性控制问题上[17-19]. 为发挥这两种控制策略的优势, 部分学者将两者有机整合, 提出基于间歇策略的事件触发机制[20-22].

2005 年, Ren 等[23] 提出了一个开放性问题: 如何设计一个分布式协议, 在使得多智能体系统达到一致性的前提下, 又能够优化某些性能指标. 针对整数阶多自主体系统, Zhang 等[24] 基于强化学习方法研究了离散时间多自主体系统的最优一致性控制问题. Zhao 等[25] 利用自适应动态规划技术, 提出了一种具有扰动的未知非线性多智能体系统的事件触发一致性跟踪控制策略. Dong 等[26] 研究了带有控制约束的连续时间系统的事件触发自适应动态规划方法. 刘晨等[27] 研究了基于事件触发策略的多自主体系统的最优主-从一致性.

相对整数阶, 分数阶微积分的分析工具不够完善, HJB 方程求解困难, 其最优一致性尚未被充分研究. 因此, 本文的主要目的就是进一步填补空白, 采用强化学习中的 actor-critic 算法研究分数阶多自主体系统的最优主-从一致性, 设计基于周期间歇事件触发策略的强化学习算法结构.

1　预备知识

分数阶微分有多种定义方式, 常用的是 Riemann-Liouville 型 (简称 R-L 型) 分数阶微分、Caputo 型分数阶微分以及 Grünwald-Letnikov 型分数阶微分等. R-L 型分数阶微分在数学上有很好的性质, 但相比而言, Caputo 型分数阶微分的初值物理意义明确, 很早就得到了广泛的应用[28]. 本文中分数阶多自主体系统的动力模型均由 Caputo 型分数阶微分描述. 下面介绍 Caputo 型分数阶微分的定义、一阶逼近以及基本性质.

定义 1[28]　Caputo 型分数阶微分算子定义:

$$ {}_{t_0}^{\mathrm{C}}\mathrm{D}_t^{\alpha} x(t) = \frac{1}{\Gamma(n-\alpha)} \int_{t_0}^{t} (t-s)^{n-\alpha-1} x^{(n)}(s)\,\mathrm{d}s, $$

其中　$\alpha > 0,\ n = [\alpha] + 1.$

根据文献 [29-31], 当$0 < \alpha < 1$, 函数$x(t) \in C^2[t_0, t_f]$时, 可得到 Caputo 型微分算子${}_{t_0}^{\mathrm{C}}\mathrm{D}_t^{\alpha} x(t)$的一阶展开式逼近:

$$ {}_{t_0}^{\mathrm{C}}\mathrm{D}_t^{\alpha} x(t) = A(\alpha)(t-t_0)^{-\alpha} x(t) + B(\alpha)(t-t_0)^{1-\alpha}\dot{x}(t) - \sum_{p=2}^{\infty} C(\alpha,p)(t-t_0)^{1-p-\alpha}\dot{x}_{(p)}(t) - \frac{x(t_0)}{\Gamma(1-\alpha)}(t-t_0)^{-\alpha}, \quad (1) $$

其中

$$ \dot{x}_{(p)}(t) = (1-p)(t-t_0)^{p-2}x(t),\ x_{(p)}(t_0) = 0, \qquad p = 2,3,\cdots, $$

$$ A(\alpha) = \frac{1}{\Gamma(1-\alpha)}\left(1 + \sum_{p=2}^{\infty}\frac{\Gamma(p-1+\alpha)}{\Gamma(\alpha)(p-1)!}\right),\ B(\alpha) = \frac{1}{\Gamma(2-\alpha)}\left(1 + \sum_{p=1}^{\infty}\frac{\Gamma(p-1+\alpha)}{\Gamma(\alpha-1)(p)!}\right), $$

$$ C(\alpha,p) = \frac{\Gamma(p-1+\alpha)}{\Gamma(2-\alpha)\Gamma(\alpha-1)(p-1)!}. $$

定义 2[28] $f(t),g(t)\in C^1[a,b],\alpha>0,\beta>0$,则

1) ${}_a^{\mathrm{C}}\mathrm{D}_t^{\alpha}(f(t)\pm g(t))={}_a^{\mathrm{C}}\mathrm{D}_t^{\alpha}f(t)\pm{}_a^{\mathrm{C}}\mathrm{D}_t^{\alpha}g(t)$;

2) ${}_a^{\mathrm{C}}\mathrm{D}_t^{\alpha}{}_a^{\mathrm{C}}\mathrm{D}_t^{-\beta}f(t)={}_a^{\mathrm{C}}\mathrm{D}_t^{\alpha-\beta}f(t)$.

2 问题描述

2.1 模型描述

考虑带有领导者的分数阶多自主体系统:

$$\begin{cases}{}_{t_0}^{\mathrm{C}}\mathrm{D}_t^{\alpha}\boldsymbol{x}_0(t)=f(t,\boldsymbol{x}_0(t)),\\ {}_{t_0}^{\mathrm{C}}\mathrm{D}_t^{\alpha}\boldsymbol{x}_i(t)=f(t,\boldsymbol{x}_i(t))+\boldsymbol{u}_i(t),\qquad i=1,2,\cdots,N,\end{cases}\tag{2}$$

其中阶数$0<\alpha<1$, $\boldsymbol{x}_0(t)=(x_{01}(t),\ x_{02}(t),\cdots,x_{0n}(t))^{\mathrm{T}}\in R^n$表示领导者的状态, $\boldsymbol{x}_i(t)=(x_{i1}(t),x_{i2}(t),\cdots,x_{in}(t))^{\mathrm{T}}\in R^n$表示第$i$个自主体的状态, $\boldsymbol{u}_i(t)$表示第i个自主体的控制输入, $f:\mathbf{R}\times R^n\to R^n$是连续可微的向量函数.

定义 3 若对任意的初始状态$\boldsymbol{x}_i(t_0)$, 可找到$\boldsymbol{u}_i(t)$使得$\lim_{t\to\infty}\|\boldsymbol{x}_i(t)-\boldsymbol{x}_0(t)\|=0$, 则称该分数阶多自主体系统 (2) 可达到主-从一致, 对$\forall i=1,2,3,\cdots,N$.

定义第i个追随者与领导者之间的状态误差如下:

$$\boldsymbol{\delta}_i(t)=\sum_{j\in N_i}a_{ij}(\boldsymbol{x}_i(t)-\boldsymbol{x}_j(t))+b_i(\boldsymbol{x}_i(t)-\boldsymbol{x}_0(t)).\tag{3}$$

将领导者和各追随者均看作节点, 得到节点集$v=\{0,1,2,\cdots,N\}$. 对称矩阵$\boldsymbol{A}=(a_{ij})_{N\times N}$, $a_{ij}\geqslant 0$表示各追随者间的通讯情况, $a_{ij}>0$表示i节点与j节点有通讯, 反之, i,j节点间无信息流通. 进而用$N_i=\{j\in v|a_{ij}\neq 0\}$来表示节点$i$的相邻节点集合. 对角矩阵$\boldsymbol{B}=(b_i)_{N\times N}$表示领导者 (0 节点) 与各追随者间的通讯情况, $b_i>0$代表 0 节点与i节点有交流, 反之没有.

则全局状态误差可表示为

$$\boldsymbol{\delta}(t)=[(\boldsymbol{L}+\boldsymbol{B})\otimes\boldsymbol{I}_n][\boldsymbol{x}(t)-\boldsymbol{x}_0(t)]=\boldsymbol{H}[\boldsymbol{x}(t)-\boldsymbol{x}_0^{\sim}(t)],\tag{4}$$

其中$\otimes$为 Kronecker 乘积符号, $\boldsymbol{x}(t)=(\boldsymbol{x}_1^{\mathrm{T}},\boldsymbol{x}_2^{\mathrm{T}},\cdots,\boldsymbol{x}_N^{\mathrm{T}})^{\mathrm{T}}\in R_N^n$表示全局状态向量, $\boldsymbol{x}_0^{\sim}(t)=(\boldsymbol{x}_0^{\mathrm{T}},\boldsymbol{x}_0^{\mathrm{T}},\cdots,\boldsymbol{x}_0^{\mathrm{T}})^{\mathrm{T}}\in R_N^n$. 定义度矩阵$\boldsymbol{D}=\mathrm{diag}\left(\sum_{j\in N_i}a_{ij}\right)_{N\times N}$, $i=1,2,\cdots,N$, 则 Laplace 矩阵$\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{A}$.

注 1 因为$\boldsymbol{H}$为正定阵, 所以$\boldsymbol{\delta}(t)\to\boldsymbol{0}$等价于$\boldsymbol{x}(t)\to\boldsymbol{x}_0^{\sim}(t)$, 即$\boldsymbol{x}_i(t)\to\boldsymbol{x}_0(t),i=1,2,\cdots,N$, 代表系统达到主-从一致.

针对分数阶多自主体系统 (2), 本文不仅考虑如何让系统达到主-从一致, 还考虑在系统达到主-从一致的过程中的能量消耗, 因此引入性能指标的概念.

定义第i个自主体的性能指标为

$$V_i(\boldsymbol{\delta}_i(t_0))=\int_{t_0}^{\infty}P(\boldsymbol{\delta}_i)+W(\boldsymbol{u}_i,\boldsymbol{u}_j)\mathrm{d}t,\tag{5}$$

其中$P(\boldsymbol{\delta}_i)=\boldsymbol{\delta}_i^{\mathrm{T}}(t)\boldsymbol{Q}_i\boldsymbol{\delta}_i(t)$是过程代价, 与一致性性能相关, 度量了系统在达到一致性过程中的一致偏差, 代表的是“运动能量”; $W(\boldsymbol{u}_i,\boldsymbol{u}_j)=\boldsymbol{u}_i^{\mathrm{T}}(t)\boldsymbol{R}_i\boldsymbol{u}_i(t)+\sum_{j\in N_i}\boldsymbol{u}_j^{\mathrm{T}}(t)\boldsymbol{R}_j\boldsymbol{u}_j(t)$是控制代价, 代表的是“控制能量”, $\boldsymbol{Q}_i\geqslant\boldsymbol{0}$, $\boldsymbol{R}_i>\boldsymbol{0},\boldsymbol{R}_j>\boldsymbol{0}$.

本文的目的是对于每个自主体i, 找到合适的控制器$\boldsymbol{u}_i(t),\boldsymbol{u}_j(t)$, 使得系统 (2) 在达到主-从一致的同时性能指标最小:

$$V_i^*(\boldsymbol{\delta}_i(t_0))=\min_{u_i}\int_{t_0}^{\infty}P(\boldsymbol{\delta}_i)+W(\boldsymbol{u}_i,\boldsymbol{u}_j)\mathrm{d}t.\tag{6}$$

由式 (6) 得自主体i的 Lyapunov 方程为

$$(V_i^{*\boldsymbol{\delta}_i})^{\mathrm{T}}\dot{\boldsymbol{\delta}}_i+P(\boldsymbol{\delta}_i)+W(\boldsymbol{u}_i^*,\boldsymbol{u}_j^*)=0.\tag{7}$$

另外, 由 Caputo 型微分算子一阶逼近式 (1) 和系统动力模型 (2) 得

$$\dot{\boldsymbol{\delta}}_i=\sum_{j\in N_i}a_{ij}(\dot{\boldsymbol{x}}_i(t)-\dot{\boldsymbol{x}}_j(t))+b_i(\dot{\boldsymbol{x}}_i(t)-\dot{\boldsymbol{x}}_0(t))=\left(b_i+\sum_{j\in N_i}a_{ij}\right)\frac{\boldsymbol{u}_i(t)}{B(\alpha)(t-t_0)^{1-\alpha}}+K(\boldsymbol{x}_0,\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{u}_j),\tag{8}$$

其中

$$K(\boldsymbol{x}_0,\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{u}_j)=\frac{b_i}{B(\alpha)(t-t_0)^{1-\alpha}}\Bigg[f(\boldsymbol{x}_i)-f(\boldsymbol{x}_0)-A(\alpha)(t-t_0)^{-\alpha}(\boldsymbol{x}_i-\boldsymbol{x}_0)+\sum_{p=2}^{\infty}C(\alpha,p)(t-t_0)^{1-p-\alpha}(\dot{\boldsymbol{x}}_{i(p)}-\dot{\boldsymbol{x}}_{0(p)})+\frac{\boldsymbol{x}_i(t_0)-\boldsymbol{x}_0(t_0)}{\Gamma(1-\alpha)}(t-t_0)^{-\alpha}\Bigg]+\frac{\sum\limits_{j\in N_i}a_{ij}}{B(\alpha)(t-t_0)^{1-\alpha}}\Bigg[f(\boldsymbol{x}_i)-f(\boldsymbol{x}_j)-\boldsymbol{u}_j-A(\alpha)(t-t_0)^{-\alpha}(\boldsymbol{x}_i-\boldsymbol{x}_j)+\sum_{p=2}^{\infty}C(\alpha,p)(t-t_0)^{1-p-\alpha}(\dot{\boldsymbol{x}}_{i(p)}-\dot{\boldsymbol{x}}_{j(p)})+\frac{\boldsymbol{x}_i(t_0)-\boldsymbol{x}_j(t_0)}{\Gamma(1-\alpha)}(t-t_0)^{-\alpha}\Bigg]\tag{9}$$

与$\boldsymbol{u}_i$无关.

则方程 (7) 等价于

$$(V_i^{*\boldsymbol{\delta}_i})^{\mathrm{T}}\left[\left(b_i+\sum_{j\in N_i}a_{ij}\right)\frac{\boldsymbol{u}_i^*}{B(\alpha)(t-t_0)^{1-\alpha}}+K(\boldsymbol{x}_0,\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{u}_j^*)\right]+P(\boldsymbol{\delta}_i)+W(\boldsymbol{u}_i^*,\boldsymbol{u}_j^*)=0.\tag{10}$$

根据 Bellman 最优性原理可得第i个自主体的最优控制为

$$\boldsymbol{u}_i^*(t)=-\boldsymbol{R}_i^{-1}\left(b_i+\sum_{j\in N_i}a_{ij}\right)\frac{V_i^{*\boldsymbol{\delta}_i}}{2B(\alpha)(t-t_0)^{1-\alpha}}.\tag{11}$$

2.2　周期间歇事件触发策略

对于分数阶多自主体系统 (2), 设计周期间歇反馈控制器:

$$\boldsymbol{u}_i(t)=\begin{cases}\boldsymbol{u}_i(t), & kT\leqslant t\leqslant(k+1-\rho)T,\\ \boldsymbol{0}, & (k+1-\rho)T\leqslant t<(k+1)T,\end{cases}\tag{12}$$

其中$0\leqslant\rho\leqslant1$为休息率, 相对地, $1-\rho$为工作率, T为控制周期, $k=0,1,2,3,\cdots$.

在周期间歇的基础上考虑集中式事件触发策略. 设第k个周期内的触发时刻集合为$\{t_1^k,t_2^k,t_3^k,\cdots,t_m^k,\cdots\}$, 则整个过程的事件触发时刻序列可表示为$\{t_1^0,t_2^0,t_3^0,\cdots,t_m^0,\cdots,t_1^k,t_2^k,t_3^k,\cdots,t_m^k,\cdots\}$. 若在第$k$个周期$[kT,(k+1)T)$上已知$t_m^k$, 则下一触发时刻$t_{m+1}^k$由下式给出:

$$t_{m+1}^k=\inf\{t>t_m^k\,|\,g(\boldsymbol{e}(t),\boldsymbol{\delta}(t),\theta,t)\geqslant0,\theta\geqslant0,t\in[kT,(k+1-\rho)T]\},\qquad\forall m,k\in\mathbf{N},\tag{13}$$

其中$\boldsymbol{e}(t)=(\boldsymbol{e}_1^{\mathrm{T}}(t),\boldsymbol{e}_2^{\mathrm{T}}(t),\cdots,\boldsymbol{e}_N^{\mathrm{T}}(t))^{\mathrm{T}}$为全局状态测量误差, $\boldsymbol{e}_i(t)=\boldsymbol{\delta}_i(t_m^k)-\boldsymbol{\delta}_i(t)$表示第$i$个自主体状态测量误差.

考虑到周期间歇事件触发策略, 自主体i的误差动力学可写为如下分段形式:

$${}^{\mathrm{C}}\mathrm{D}^{\alpha}\boldsymbol{\delta}_i(t)=\begin{cases}\displaystyle\sum_{j\in N_i}a_{ij}(f(t,\boldsymbol{x}_i(t))-f(t,\boldsymbol{x}_j(t))+\boldsymbol{u}_i(t_m^k)-\boldsymbol{u}_j(t_m^k))+b_i(f(t,\boldsymbol{x}_i(t))-f(t,\boldsymbol{x}_0(t))+\boldsymbol{u}_i(t_m^k)),\\ \qquad\qquad t\in[t_m^k,t_{m+1}^k)\cap[kT,(k+1-\rho)T],\\ \displaystyle\sum_{j\in N_i}a_{ij}(f(t,\boldsymbol{x}_i(t))-f(t,\boldsymbol{x}_j(t)))+b_i(f(t,\boldsymbol{x}_i(t))-f(t,\boldsymbol{x}_0(t))),\qquad t\in((k+1-\rho)T,(k+1)T),\end{cases}$$

其中$\boldsymbol{u}_i(t_m^k)$表示(t_m^k,t_{m+1}^k)区间内i自主体的控制输入.

注 2　式 (13) 中事件触发条件$g(\boldsymbol{e}(t),\boldsymbol{\delta}(t),\theta,t)$可根据具体一致性种类和控制策略来设计. 针对分数阶多自主体系统的事件触发条件大致可分为三类: 依赖于状态[26]、依赖于指数函数[16]、依赖于状态和指数函数的混合[20].

注 3　周期间歇事件触发策略仅在工作区间$[kT,(k+1-\rho)T]$, $k\in\mathbf{N}$内采用事件触发策略, 在其他时间段不对系统施加控制. 当$\rho=0$时, 此策略退化为事件触发控制策略; 当$\rho=1$时, 此策略退化为事件触发脉冲控制策略.

3　基于 actor-critic 算法的近似最优控制

Actor-critic 算法是强化学习中的一种算法, 简要原理是 actor 来做动作, critic 对 actor 做出的动作给予评价. 评价分为奖励、惩罚两种. actor 通过得到的评价不断调整自己的动作以得到更多的奖励. 下面用 critic 网络拟合性能指标函数, actor 网络拟合控制器$\boldsymbol{u}_i(t)$. 算法整体框架详见文后附录.

3.1 Critic 网络设计

根据式 (5), 确定 critic 网络的输入$\boldsymbol{Z}_{ci}(t)$必须包含$\boldsymbol{\delta}_i(t),\hat{\boldsymbol{u}}_i(t),\hat{\boldsymbol{u}}_j(t)(j\in N_i)$的信息($\hat{\boldsymbol{u}}_i(t),\hat{\boldsymbol{u}}_j(t)$由 actor 网络生成). 对于第$i$个自主体, 网络拟合的性能指标为

$$\hat{V}_i(t)=\boldsymbol{W}_{ci}(t)\psi_{\rm c}(\boldsymbol{Y}_{ci}(t)\boldsymbol{Z}_{ci}(t)), \tag{14}$$

其中$\boldsymbol{Y}_{ci}(t)$表示输入层到隐含层的权重, $\boldsymbol{W}_{ci}(t)$表示隐含层到输出层的权重, $\psi_{\rm c}(\cdot)$为激活函数.

由式 (7) 可得

$$(\hat{V}_i^{\boldsymbol{\delta}_i}(t))^{\rm T}\dot{\boldsymbol{\delta}}_i(t)=-P(\boldsymbol{\delta}_i(t))-W(\hat{\boldsymbol{u}}_i(t)), \tag{15}$$

进而

$$\frac{\hat{V}_i(t^+)-\hat{V}_i(t)}{t^+-t}\approx -P(\boldsymbol{\delta}_i(t))-W(\hat{\boldsymbol{u}}_i(t)). \tag{16}$$

因为网络拟合存在重构误差, 所以定义 critic 网络的误差函数:

$$e_{ci}(t)=\hat{V}_i(t)-\hat{V}_i(t^-)+(t-t^-)[P(\boldsymbol{\delta}_i(t^-))+W(\hat{\boldsymbol{u}}_i(t^-))]. \tag{17}$$

Critic 网络训练的目的为: 选择合适的$\boldsymbol{Y}_{ci}(t),\boldsymbol{W}_{ci}(t)$使得$E_{ci}(t)=\dfrac{1}{2}e_{ci}^2(t)$尽量小.

当达到周期间歇事件触发阈值时, 使用梯度下降法对网络权重进行更新, 否则权重不更新, 具体更新方式如下:

$$\begin{cases}\dot{\boldsymbol{W}}_{ci}(t)=\boldsymbol{0}, & t\in(t_m^k,t_{m+1}^k)\cup((k+1-\rho)T,(k+1)T),\\ \boldsymbol{W}_{ci}^+(t)=\boldsymbol{W}_{ci}(t)-\beta_{ci}\dfrac{\partial E_{ci}(t)}{\partial \boldsymbol{W}_{ci}(t)}=\boldsymbol{W}_{ci}(t)-\beta_{ci}\dfrac{\partial E_{ci}(t)}{\partial e_{ci}(t)}\dfrac{\partial e_{ci}(t)}{\partial \hat{V}_i(t)}\dfrac{\partial \hat{V}_i(t)}{\partial \boldsymbol{W}_{ci}(t)}, & \\ & t\in\{t_1^0,t_2^0,t_3^0,\cdots,t_m^0,\cdots,t_1^k,t_2^k,t_3^k,\cdots,t_m^k,\cdots\},\end{cases} \tag{18}$$

$$\begin{cases}\dot{\boldsymbol{Y}}_{ci}(t)=\boldsymbol{0}, & t\in(t_m^k,t_{m+1}^k)\cup((k+1-\rho)T,(k+1)T),\\ \boldsymbol{Y}_{ci}^+(t)=\boldsymbol{Y}_{ci}(t)-\beta_{ci}\dfrac{\partial E_{ci}(t)}{\partial \boldsymbol{Y}_{ci}(t)}=\boldsymbol{Y}_{ci}(t)-\beta_{ci}\dfrac{\partial E_{ci}(t)}{\partial e_{ci}(t)}\dfrac{\partial e_{ci}(t)}{\partial \hat{V}_i(t)}\dfrac{\partial \hat{V}_i(t)}{\partial \psi_{\rm c}(\boldsymbol{Y}_{ci}(t)\boldsymbol{Z}_{ci}(t))}\dfrac{\partial \psi_{\rm c}(\boldsymbol{Y}_{ci}(t)\boldsymbol{Z}_{ci}(t))}{\partial \boldsymbol{Y}_{ci}(t)}, & \\ & t\in\{t_1^0,t_2^0,t_3^0,\cdots,t_m^0,\cdots,t_1^k,t_2^k,t_3^k,\cdots,t_m^k,\cdots\},\end{cases} \tag{19}$$

其中β_{ci}为学习率.

3.2 Actor 网络设计

与 critic 网络类似, actor 网络同样采用三层的网络结构. 对于第i个自主体, 以$\boldsymbol{\delta}_i(t)$作为 actor 网络的输入, 得到网络拟合的控制器为

$$\hat{\boldsymbol{u}}_i(t)=\boldsymbol{W}_{ai}(t)\psi_{\rm a}(\boldsymbol{Y}_{ai}(t)\boldsymbol{\delta}_i(t)), \tag{20}$$

其中$\boldsymbol{Y}_{ai}(t)$表示输入层到隐含层的权重, $\boldsymbol{W}_{ai}(t)$表示隐含层到输出层的权重, $\psi_{\rm a}(\cdot)$为激活函数.

无论是 critic 网络还是 actor 网络, 最终目标是找到合适的控制器$\hat{\boldsymbol{u}}_i(t)$使得系统达到主-从一致时性能指标$\hat{V}_i(t)$最小(理想目标是$U_{\rm c}=0$), 所以定义 actor 网络的误差函数为

$$e_{ai}(t)=\hat{V}_i(t)-U_{\rm c}=\hat{V}_i(t). \tag{21}$$

Actor 网络训练的目的为: 选择合适的$\boldsymbol{Y}_{ai}(t),\boldsymbol{W}_{ai}(t)$使得$E_{ai}(t)=\dfrac{1}{2}e_{ai}^2(t)$尽量小.

Actor 网络的权值更新方法与 critic 网络类似, 具体公式如下:

$$\begin{cases}\dot{\boldsymbol{W}}_{ai}(t)=\boldsymbol{0}, & t\in(t_m^k,t_{m+1}^k)\cup((k+1-\rho)T,(k+1)T),\\ \boldsymbol{W}_{ai}^+(t)=\boldsymbol{W}_{ai}(t)-\beta_{ai}\dfrac{\partial E_{ai}(t)}{\partial e_{ai}(t)}\dfrac{\partial e_{ai}(t)}{\partial \hat{V}(t)}\dfrac{\partial \hat{V}(t)}{\partial \psi_{\rm c}(\boldsymbol{Y}_{ci}\boldsymbol{Z}_{ci}(t))}\dfrac{\partial \psi_{\rm c}(\boldsymbol{Y}_{ci}\boldsymbol{Z}_{ci}(t))}{\partial \boldsymbol{Z}_{ci}(t)}\dfrac{\partial \boldsymbol{Z}_{ci}(t)}{\partial \hat{\boldsymbol{u}}_i(t)}\dfrac{\partial \hat{\boldsymbol{u}}_i(t)}{\partial \boldsymbol{W}_{ai}(t)}, & \\ & t\in\{t_1^0,t_2^0,t_3^0,\cdots,t_m^0,\cdots,t_1^k,t_2^k,t_3^k,\cdots,t_m^k,\cdots\},\end{cases} \tag{22}$$

$$\begin{cases}\dot{\boldsymbol{Y}}_{ai}(t)=\boldsymbol{0}, & t\in(t_m^k,t_{m+1}^k)\cup((k+1-\rho)T,(k+1)T),\\ \boldsymbol{Y}_{ai}^+(t)=\boldsymbol{Y}_{ai}(t)-\beta_{ai}\dfrac{\partial E_{ai}(t)}{\partial e_{ai}(t)}\dfrac{\partial e_{ai}(t)}{\partial \hat{V}(t)}\dfrac{\partial \hat{V}(t)}{\partial \psi_{\rm c}(\boldsymbol{Y}_{ci}\boldsymbol{Z}_{ci}(t))}\dfrac{\partial \psi_{\rm c}(\boldsymbol{Y}_{ci}\boldsymbol{Z}_{ci}(t))}{\partial \hat{\boldsymbol{u}}_i(t)}\dfrac{\partial \hat{\boldsymbol{u}}_i(t)}{\partial \boldsymbol{Y}_{ai}(t)}, & \\ & t\in\{t_1^0,t_2^0,t_3^0,\cdots,t_m^0,\cdots,t_1^k,t_2^k,t_3^k,\cdots,t_m^k,\cdots\},\end{cases} \tag{23}$$

其中β_{ai}为学习率.

注 4　本文将分数阶微分的一阶导近似展开式 (1) 和文献 [27] 中整数阶多自主体系统的事件触发自适应动态规划算法有机整合, 进一步考虑了间歇策略, 针对分数阶多自主体系统的最优主-从一致性, 设计了基于周期间歇事件触发的强化学习算法.

4　数值仿真

例 1　考虑带有 1 个领导者, 3 个追随者的分数阶多自主体系统, 网络拓扑图如图 1.

选取$\alpha=0.95$, $\boldsymbol{A}=[0\ 1\ 0;1\ 0\ 1;0\ 1\ 0]$, $\boldsymbol{B}=[1\ 0\ 0;0\ 0\ 0;0\ 0\ 0]$ $f(x_i)=-2\sin(x_i)+\tanh(x_i), i=0,1,2,3$, 初始状态$x_0(0)=5, x_1(0)=-3, x_2(0)=-1, x_3(0)=2.8$, 时间步长$h=0.001$ s. 若无任何控制器作用, 各自主体的轨迹如图 2.

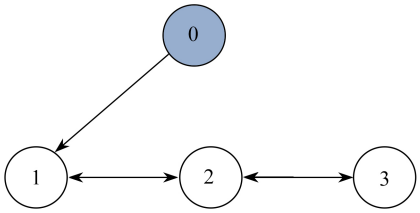

图 1　多自主体系统网络拓扑图 (1 个领导者, 3 个追随者)

Fig. 1　The net topology of the multi-agent system (1 leader, 3 followers)

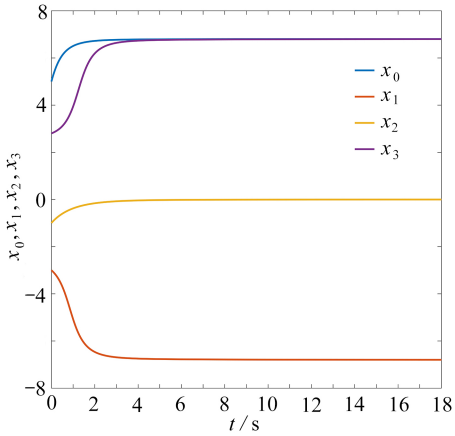

图 2　无控制器作用时, 各自主体的状态轨迹 (1 个领导者, 3 个追随者)

Fig. 2　State trajectories of each agent without controllers (1 leader, 3 followers)

设置基于周期间歇的事件触发策略: $T=3.5$ s, $\rho\approx0.143$, $\boldsymbol{g}(t)=\boldsymbol{e}(t)-\theta\boldsymbol{\delta}(t)$, 取$\theta=0.06$, 权值矩阵的初值在区间 [−0.025,0.025] 中随机选取, 并归一化处理, 其他网络参数设置如表 1.

表 1　网络参数设置

Table 1　Values of networks' parameters

parameter	meaning	value
β_{c1}	learning rate of the critic network	0.1
β_{a1}	learning rate of the actor network	0.1
$T_{c,error}$	threshold for the critic network	10^{-10}
$T_{a,error}$	threshold for the actor network	10^{-10}
N_{c1}	number of hidden nodes in the critic network	5
N_{a1}	number of hidden nodes in the critic network	3
$\psi_c(\cdot)$	activation function of the critic network	$\tanh(\cdot)$
$\psi_a(\cdot)$	activation function of the actor network	$\tanh(\cdot)$

在该策略控制作用下的数值仿真结果如图 3~5 所示. 图 3 为各自主体的状态轨迹图, 表示系统约在 10 s 达到主-从一致的状态. 图 4 为全局状态测量误差$\|\boldsymbol{e}(t)\|$及事件触发阈值的变化曲线, 可看出在接近 9 s 的时候$\|\boldsymbol{e}(t)\|$便趋于 0. 图 5 为基于周期间歇的事件触发时刻图, 描述了在 0~18 s 中事件触发时刻的具体分布: 0~3 s, 3.5~6.5 s, 7~10 s, 10.5~13.5 s, 14~17 s, 17.5~18 s 为控制器工作时间; 3~3.5 s, 6.5~7 s, 10~10.5 s, 13.5~14 s, 17~17.5 s 为控制器休息时间.

注 5　本文将间歇的事件触发机制有机整合起来, 研究了分数阶多自主体系统的最优主-从一致性. 目前该方向仅有少量成果. 文献 [20] 采用了间歇事件触发策略, 对分数阶多自主体系统进行了有界性分析, 对于一致性的研究尚未有文献涉及.

例 2　考虑带有 1 个领导者, 4 个追随者的分数阶多自主体系统, 拓扑结构如图 6.

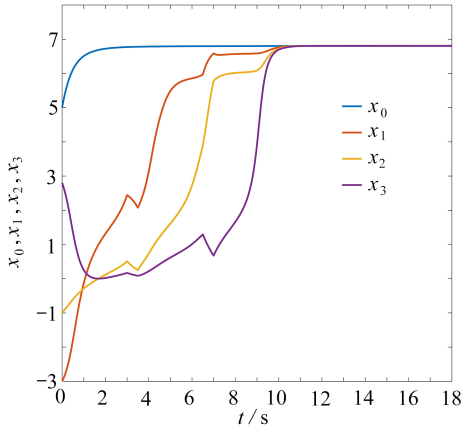

图 3 各自主体的状态轨迹 (1 个领导者, 3 个追随者)

Fig. 3 State trajectories of each agent (1 leader, 3 followers)

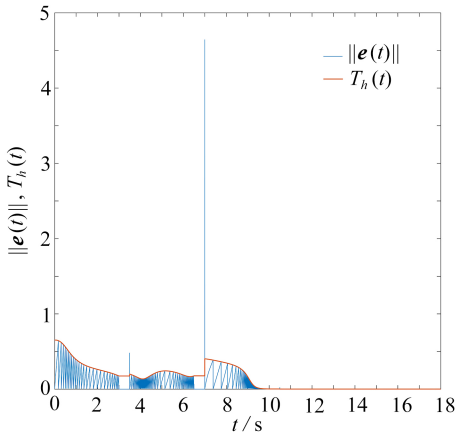

图 4 $\|\boldsymbol{e}(t)\|$及触发阈值变化曲线 (1 个领导者, 3 个追随者)

Fig. 4 The error and the trigger threshold (1 leader, 3 followers)

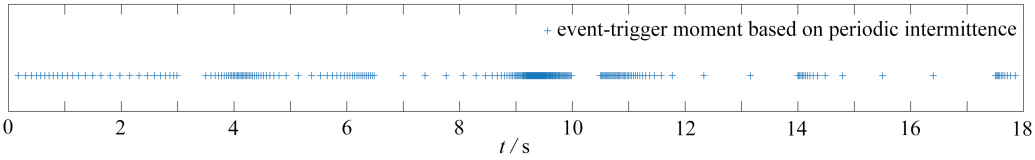

图 5 周期间歇事件触发时刻分布

Fig. 5 The event-trigger moment distribution of periodic intermittence

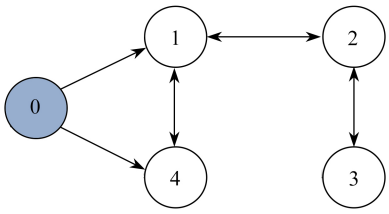

图 6 多自主体系统网络拓扑图 (1 个领导者, 4 个追随者)

Fig. 6 The net topology of the multi-agent system (1 leader, 4 followers)

选取α=0.86, $\boldsymbol{A}$=[0 1 0 1;1 0 1 0;0 1 0 0;1 0 0 0], $\boldsymbol{B}$=[1 0 0 0;0 0 0 0;0 0 0 0;0 0 0 1], $f(x_i)$=tanh(0.01x_i)−2cos(x_i), $i=0,1,2,3,4$, 初始状态$x_0(0)=5, x_1(0)=4, x_2(0)=3, x_3(0)=2, x_4(0)=6$,时间步长$h=0.001$ s.若无任何控制器作用, 各自主体的轨迹如图 7 所示.

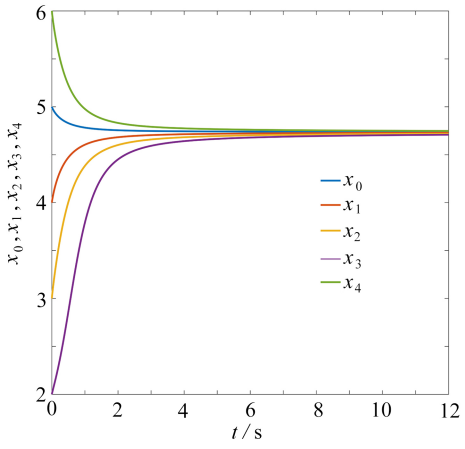

图 7 无控制器作用时, 各自主体的状态轨迹 (1 个领导者, 4 个追随者)

Fig. 7 State trajectories of each agent without controllers (1 leader, 4 followers)

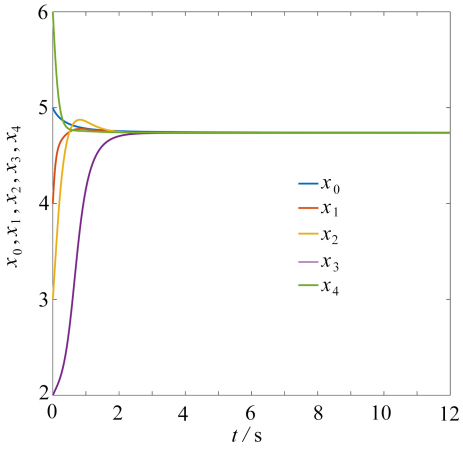

图 8 各自主体的状态轨迹 (1 个领导者, 4 个追随者)

Fig. 8 State trajectories of each agent (1 leader, 4 followers)

设置基于周期间歇的事件触发策略: $\rho=0$, $g(t)=\|\boldsymbol{e}(t)\|-\mathrm{e}^{-0.5\theta t}$, 即事件触发策略. 选取参数$\theta=1.9$, 其他网络参数如同例 1. 数值仿真结果如图 8~10 所示. 图 8 为本文所设计控制器作用下各自主体的状态轨迹图. 由图 8 看出, 系统在不到 3 s 的时间内就达到了主-从一致. 图 9 为全局状态测量误差$\|\boldsymbol{e}(t)\|$及事件触发阈值的变化曲线, 其表明系统误差在慢慢变小, 并在 3 s 后非常接近于 0. 图 10 为事件触发时刻图, 描述了 0~12 s 内事件触发的具体时刻分布, 触发 40 次.

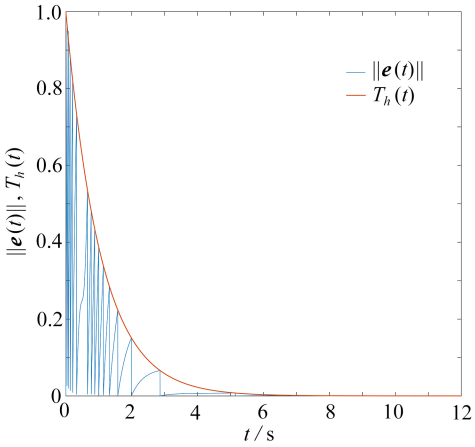

图 9　$\|\boldsymbol{e}(t)\|$及触发阈值变化曲线 (1 个领导者, 4 个追随者)

Fig. 9　The error and the trigger threshold (1 leader, 4 followers)

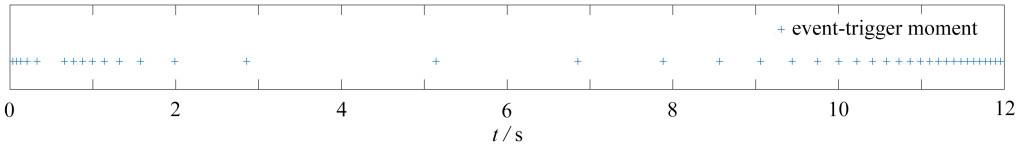

图 10　事件触发时刻分布

Fig. 10　The event-trigger moment distribution

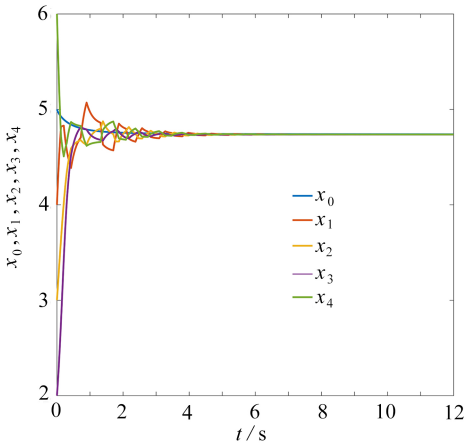

图 11　文献 [16] 控制器下, 各自主体的状态轨迹图

Fig. 11　State trajectories of each agent under ref. [16]

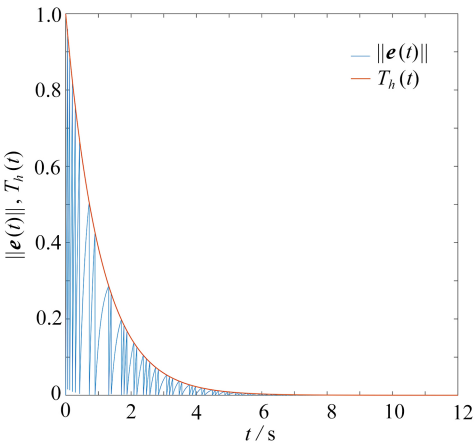

图 12　$\|\boldsymbol{e}(t)\|$及触发阈值变化曲线

Fig. 12　The error $\|\boldsymbol{e}(t)\|$ and the trigger threshold

注 6　图 11 展示了文献 [16] 中控制器作用下各自主体的状态轨迹. 对比图 8 和图 11, 网络拟合控制器将系统趋于一致的速度提高了不止 1 s. 图 12 为文献 [16] 控制器下系统达到主-从一致过程中的全局状态测量误差变化情况. 图 13 为事件触发时刻图, 描述了 0~12 s 内事件触发的具体时刻分布, 触发 104 次. 通过图 10 和图 13 可明显看出, 在系统达到主-从一致的过程中本文所设计控制器作用下的事件触发次数较少, 一定程度上减少了通讯成本.

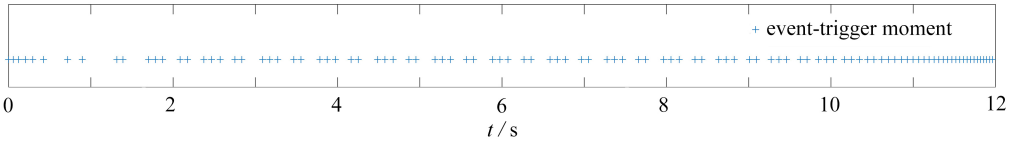

图 13 事件触发时刻分布

Fig. 13 The event-trigger moment distribution

5 总 结

本文借助分数阶微分的一阶近似逼近和强化学习中的 actor-critic 算法, 研究了在控制器周期间歇时, 分数阶多自主体系统在事件触发策略下的最优主-从一致性问题, 最终设计出基于 actor-critic 算法的控制策略, 并通过仿真验证了其有效性.

附 录

Actor-critic 近似最优控制算法整体框架如下:

输入: actor 模型$\pi_{\boldsymbol{W}_{\mathrm{a}i},\boldsymbol{Y}_{\mathrm{a}i}}(\boldsymbol{\delta}_i)$, critic 模型$V_{\boldsymbol{W}_{\mathrm{c}i},\boldsymbol{Y}_{\mathrm{c}i}}(\boldsymbol{x}_i,\boldsymbol{u}_i,\boldsymbol{u}_j)$, $i=1,2,\cdots,N$.

For $i=1,2,\cdots,N$

初始化状态$\boldsymbol{x}_i$, 得到初始$\boldsymbol{\delta}_i$, 初始化参数$\boldsymbol{W}_{\mathrm{a}i},\boldsymbol{Y}_{\mathrm{a}i},\boldsymbol{W}_{\mathrm{c}i},\boldsymbol{Y}_{\mathrm{c}i}$

End for

For $t\in[kT,(k+1)T)$

If $t\in[kT,(k+1-\rho)T]$, 即控制器处于工作时间

For $i=1,2,\cdots,N$

遵循策略$\pi_{\boldsymbol{W}_{\mathrm{a}i},\boldsymbol{Y}_{\mathrm{a}i}}(\boldsymbol{\delta}_i)$, 得到控制$\boldsymbol{u}_i$

在$\boldsymbol{u}_i$的作用下, 自主体i得到新状态$\boldsymbol{x}_i'$以及回报$r_i=V_i$

End for

计算全局状态误差$\boldsymbol{\delta}(t)=(\boldsymbol{\delta}_1^{\mathrm{T}}(t),\boldsymbol{\delta}_2^{\mathrm{T}}(t),\cdots,\boldsymbol{\delta}_N^{\mathrm{T}}(t))^{\mathrm{T}}$

全局状态测量误差$\boldsymbol{e}(t)=(\boldsymbol{e}_1^{\mathrm{T}}(t),\boldsymbol{e}_2^{\mathrm{T}}(t),\cdots,\boldsymbol{e}_N^{\mathrm{T}}(t))^{\mathrm{T}}$

If 系统达到事件触发条件的阈值: $g(\boldsymbol{e}(t),\boldsymbol{\delta}(t),\theta,\beta,t)\geqslant 0$

For $i=1,2,\cdots,N$

If 自主体i的神经网络满足网络误差阈值

更新网络权重, 得到新的策略$\pi'_{\boldsymbol{W}_{\mathrm{a}i},\boldsymbol{Y}_{\mathrm{a}i}}(\boldsymbol{\delta}_i)$

$\boldsymbol{W}_{\mathrm{a}i}\leftarrow\boldsymbol{W}_{\mathrm{a}i}+\boldsymbol{\beta}_{\mathrm{a}i}\nabla_{\mathrm{a}i}\pi_{\boldsymbol{W}_{\mathrm{a}i},\boldsymbol{Y}_{\mathrm{a}i}}(\boldsymbol{\delta}_i)$

$\boldsymbol{Y}_{\mathrm{a}i}\leftarrow\boldsymbol{Y}_{\mathrm{a}i}+\boldsymbol{\beta}_{\mathrm{a}i}\nabla_{\mathrm{a}i}\pi_{\boldsymbol{W}_{\mathrm{a}i},\boldsymbol{Y}_{\mathrm{a}i}}(\boldsymbol{\delta}_i)$

$\boldsymbol{W}_{\mathrm{c}i}\leftarrow\boldsymbol{W}_{\mathrm{c}i}+\boldsymbol{\beta}_{\mathrm{c}i}\nabla_{\mathrm{c}i}V_{\boldsymbol{W}_{\mathrm{c}i},\boldsymbol{Y}_{\mathrm{c}i}}(\boldsymbol{x}_i,\boldsymbol{u}_i,\boldsymbol{u}_j)$

$\boldsymbol{Y}_{\mathrm{c}i}\leftarrow\boldsymbol{Y}_{\mathrm{c}i}+\boldsymbol{\beta}_{\mathrm{c}i}\nabla_{\mathrm{c}i}V_{\boldsymbol{W}_{\mathrm{c}i},\boldsymbol{Y}_{\mathrm{c}i}}(\boldsymbol{x}_i,\boldsymbol{u}_i,\boldsymbol{u}_j)$

End if

End for

End if

Else, 即控制器处于休息时间

For $i=1,2,\cdots,N$

根据$\boldsymbol{u}_i=0$时的状态方程计算得出自主体i的新状态$\boldsymbol{x}_i'$

End for

End if

End for

输出: 最优控制策略$\pi^*_{\boldsymbol{W}_{ai},\boldsymbol{Y}_{ai}}(\boldsymbol{\delta}_i)$, $i=1,2,\cdots,N$.

参考文献(References):

[1] CORTÉS J, BULLO F. Coordination and geometric optimization via distributed dynamical systems[J]. *SIAM Journal on Control and Optimization*, 2005, **44**(5): 1543-1574.

[2] FAX J A, MURRAY R M. Information flow and cooperative control of vehicle formations[J]. *IEEE Transactions on Automatic Control*, 2004, **49**(9): 1465-1476.

[3] YU W W, CHEN G R, WANG Z D, et al. Distributed consensus filtering in sensor networks[J]. *IEEE Transactions on Systems, Man, and Cybernetics* (*Part B*): *Cybernetics*, 2009, **39**(6): 1568-1577.

[4] BEARD R W, MCLAIN T W, GOODRICH M A, et al. Coordinated target assignment and intercept for unmanned air vehicles[J]. *IEEE Transactions on Robotics and Automation*, 2002, **18**(6): 911-922.

[5] FAX J A, MURRAY R M. Information flow and cooperative control of vehicle formations[J]. *IFAC Proceedings Volumes*, 2002, **35**(1): 115-120.

[6] YU W W, WANG H, CHENG F, et al. Second-order consensus in multiagent systems via distributed sliding mode control[J]. *IEEE Transactions on Cybernetics*, 2017, **47**(8): 1872-1881.

[7] YU W W, CHEN G R, CAO M, et al. Second-order consensus for multiagent systems with directed topologies and nonlinear dynamics[J]. *IEEE Transactions on Systems, Man, and Cybernetics* (*Part B*): *Cybernetics*, 2010, **40**(3): 881-891.

[8] WEN G H, YU W W, XIA Y Q, et al. Distributed tracking of nonlinear multiagent systems under directed switching topology: an observer-based protocol[J]. *IEEE Transactions on Systems, Man, and Cybernetics:Systems*, 2017, **47**(5): 869-881.

[9] WEN G H, YU W W, LI Z H, et al. Neuro-adaptive consensus tracking of multiagent systems with a high-dimensional leader[J]. *IEEE Transactions on Cybernetics*, 2017, **47**(7): 1730-1742.

[10] SUN W, LI Y, LI C P, et al. Convergence speed of a fractional order consensus algorithm over undirected scale-free networks[J]. *Asian Journal of Control*, 2011, **13**(6): 936-946.

[11] CHAO S, CAO J D. Consensus of fractional-order linear systems[C]//*2013 9th Asian Control Conference (ASCC)*. Istanbul, Turkey, 2013.

[12] YU W W, LI Y, WEN G H, et al. Observer design for tracking consensus in second-order multi-agent systems: fractional order less than two[J]. *IEEE Transactions on Automatic Control*, 2017, **62**(2): 894-900.

[13] ASTROM K J, BERNHARDSSON B. Comparison of periodic and event based sampling for first-order stochastic systems[C]//*14th IFAC World Congress*. Beijing, China, 1999.

[14] DIMAROGONAS D V, JOHANSSON K H. Event-triggered control for multi-agent systems[C]//*Proceedings of the 48th IEEE Conference on Decision and Control, CDC 2009, Combined With the 28th Chinese Control Conference*. Shanghai, China, 2009.

[15] XU G H, CHI M, HE D X, et al. Fractional-order consensus of multi-agent systems with event-triggered control[C]//*2014 11th IEEE International Conference on Control & Automation* (*ICCA*). Taichung, 2014.

[16] WANG F, YANG Y Q. Leader-following consensus of nonlinear fractional-order multi-agent systems via event-triggered control[J]. *International Journal of Systems Science*, 2017, **48**(3): 571-577.

[17] YE Y Y, SU H S. Consensus of delayed fractional-order multiagent systems with intermittent sampled data[J]. *IEEE Transactions on Industrial Informatics*, 2019, **16**(6): 3828-3837.

[18] XU L G, LIU W, HU H X, et al. Exponential ultimate boundedness of fractional-order differential systems via periodically intermittent control[J]. *Nonlinear Dynamics*, 2019, **96**(2): 1665-1675.

[19] XU Y, LI Q, LI W X. Periodically intermittent discrete observation control for synchronization of fractional-order coupled systems[J]. *Communications in Nonlinear Science and Numerical Simulation*, 2019, **74**: 219-235.

[20] CHANG Q, HU A H, YANG Y Q, et al. Pinning exponential boundedness of fractional-order multi-agent systems with intermittent combination event-triggered protocol[J]. *International Journal of Systems Science*, 2020, **52**(4): 874-888.

[21] HU A, HP JU, HU M. Consensus of nonlinear multiagent systems with intermittent dynamic event-triggered protocols[J]. *Nonlinear Dynamics*, 2021, **104**: 1299-1313.

[22] LIU X Y, FU H B, LIU L. Leader-following mean square consensus of stochastic multi-agent systems via periodically intermittent event-triggered control[J]. *Neural Processing Letters*, 2020, **53**(1): 275-298.

[23] REN W, BEARD R W, ATKINS E M. A survey of consensus problems in multi-agent coordination[C]//*Proceedings of the 2005, American Control Conference.* Portland, OR, USA, 2005.

[24] ZHANG H G, JIANG H, LUO Y H, et al. Data-driven optimal consensus control for discrete-time multi-agent systems with unknown dynamics using reinforcement learning method[J]. *IEEE Transactions on Industrial Electronics*, 2017, **64**(5): 4091-4100.

[25] ZHAO W, YU W W, ZHANG H P. Event-triggered optimal consensus tracking control for multi-agent systems with unknown internal states and disturbances[J]. *Nonlinear Analysis Hybrid Systems*, 2019, **33**: 227-248.

[26] DONG L, ZHONG X N, SUN C Y, et al. Event-triggered adaptive dynamic programming for continuous-time systems with control constraints[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, **28**(8): 1941-1952.

[27] 刘晨, 刘磊. 基于事件触发策略的多智能体系统的最优主-从一致性分析[J]. 应用数学和力学, 2019, **40**(11): 1278-1288. (LIU Chen, LIU Lei. Optimal leader-follower consensus of multi-agent systems based on the event-triggered strategy[J]. *Applied Mathematics and Mechanics*, 2019, **40**(11): 1278-1288.(in Chinese))

[28] PODLUBNY I. *Fractional Differential Equations*[M]. New York, USA: Academic, 1999.

[29] ATANACKOVIC T M, STANKOVIC B. On a numerical scheme for solving differential equations of fractional order[J]. *Mechanics Research Communications*, 2008, **35**(7): 429-438.

[30] POOSEH S, ALMEIDA R, TORRES D. Fractional order optimal control problems with free terminal time[J]. *Journal of Industrial & Management Optimization*, 2014, **10**(2): 363-381.

[31] ILBAS A A A, SRIVASTAVA H M, TRUJILLO J J. *Theory and Applications of Fractional Differential Equations*[M]. *North-Holland Mathematics Studies*, **204**. Elsevier, 2006.