

未知异构多智能体系统无模型自适应动态规划同步控制

夏丽娜¹, 李擎¹, 宋睿卓¹, 王子涵¹, 许镇²

(1. 北京科技大学自动化学院, 北京 100083;

2. 北京科技大学城镇化与城市安全研究院, 北京 100083)

摘要: 多智能体协同在众多领域逐渐得到应用, 但是还有很多仍未解决的问题。为此, 研究了基于无模型自适应动态规划 (MFADP) 的未知异构多智能体系统同步控制问题, 其中领航者智能体和追随者智能体动态模型均是未知的。首先为每一个追随者智能体设计观测器来估计领航者智能体的信息, 包括领航者智能体状态以及领航者智能体系统动态矩阵。然后利用追随者智能体动态模型信息和观测器动态模型信息构建增广系统, 运用贝尔曼最优性原理求得最优控制器以及对应的代数黎卡提方程。在追随者智能体系统动态模型信息未知的情况下, 引出了 MFADP 算法。最后使用双质量-弹簧系统来验证所提算法的有效性。

关键词: 多智能体系统; 代数黎卡提方程; 基于无模型自适应动态规划

中图分类号: TP18

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202143

Synchronization control of unknown heterogeneous multi-agent system via model-free adaptive dynamic programming

XIA Lina¹, LI Qing¹, SONG Ruizhuo¹, WANG Zihan¹, XU Zhen²

1. School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

2. Research Institute of Urbanization and Urban Safety, University of Science and Technology Beijing, Beijing 100083, China

Abstract: Synchronization of multi-agent system has been gradually applied in the most fields, but there are still many unsolved problems. The synchronization control of unknown heterogeneous multi-agent system based on model-free adaptive dynamic programming (MFADP) algorithm was studied. Firstly, an observer was designed for each follower to estimate the information of the leader, including the state and the system dynamic matrix of the leader. Then, the optimal controller was obtained by exploiting the Bellman optimality principle. Under the condition that the dynamics of the follower was unknown, a MFADP algorithm was proposed. Finally, two-mass-spring systems were used to verify the effectiveness of the algorithm.

Key words: multi-agent system, algebraic Riccati equation, model-free adaptive dynamic programming

1 引言

随着科学技术的飞速发展及生产规模的不断扩大, 社会上存在很多大型、复杂的现实问题, 而解决这类问题已超出了单个智能体的能力。随着通

信技术、计算机技术、机器人技术、传感器以及执行器的小型化发展, 多个个体协同完成任务成为可能, 并且逐渐在众多领域得到应用。多智能体系统 (multi-agent system, MAS) 同步控制吸引了很多学者的注意, 已经被成功地应用于各种学科^[1]。根据智

收稿日期: 2021-09-22; 修回日期: 2021-11-09

通信作者: 宋睿卓, ruizhuosong@ustb.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61873300, No.61722312); 中央高校基本科研业务费专项资金项目 (No.FRF-MP-20-11, No.FRF-IDRY-20-030)

Foundation Items: The National Natural Science Foundation of China (No.61873300, No.61722312), The Fundamental Research Funds for the Central Universities (No.FRF-MP-20-11, No.FRF-IDRY-20-030)

能体的动态模型和维度, MAS 被分为同构 MAS 和异构 MAS。根据领航者智能体(以下简称领航者)的数量, MAS 同步控制可以分为无领航者问题、领航者-追随者问题以及包含问题^[2]。然而,在大多数实际应用下,智能体的动态模型和维度可能是不相同的,例如某些协作任务中参与的智能体可能包括地面无人机、水下无人机和空中无人机。因此本文研究了异构 MAS 领航者-追随者问题。在 MAS 通信网络中,并不是所有的追随者智能体(以下简称追随者)都能够获取领航者的信息,如领航者状态以及系统动态矩阵信息。Yang Y L 等人^[3]和 Li Q 等人^[4]假设领航者系统动态矩阵已知,设计观测器来估计领航者的状态。在实际应用中,多智能体系统动态矩阵以及状态信息可能均是未知的。因此,未知领航者状态以及系统动态矩阵信息下的异构 MAS 同步问题值得进一步研究。

自适应动态规划(adaptive dynamic programming, ADP)是在动态规划的基础上发展而来的,用于解决动态规划中的维数灾难问题。为了方便 ADP 在线运行,韦伯(Werbos)给出了两种 ADP 基本结构,其中一种被称为启发式动态规划,另一种被称为二次启发式规划。进一步地,韦伯又提出了两种改进结构:控制依赖启发式动态规划和控制依赖二次启发式规划。这两种结构与前两种结构的主要区别是:评价网的输入不再只有系统状态,还包含控制输入。考虑系统动态是否已知,ADP 又分为基于模型的自适应动态规划(model-based adaptive dynamic programming, MBADP)和无模型自适应动态规划(model-free adaptive dynamic programming, MFADP)。Modares H 等人^[5]针对系统动态完全未知的非线性单智能体连续时间系统设计了一种最优跟踪控制器,无须系统的动态信息,并且证明了该方法的收敛性。但是该方法只针对单智能体系统,其结果无法扩展到 MAS 中。刘莹莹等人^[6]研究了离散异构 MAS 的输出同步问题,提出了一种基于数据的 ADP 方法。之后,Modares H 等人^[7]和 Xia L N 等人^[8]使用 MFADP 解决了异构连续时间非线性 MAS 的最优输出同步问题和带非对称输入受限的异构 MAS 最优输出同步问题,但是均需要领航者的系统动态矩阵信息。于是,本文研究了基于 MFADP 未知异构连续时间线性 MAS 同步控制问题,不需要考虑系统动态模型信息,即领航者和追随者系统动态模型信息均是未知的。

2 领航者-追随者一致性最优问题

本文研究有向图 $G(V, \mathfrak{I}, A)$, 其中有向图的三要素为节点集 $V = \{1, \dots, N\}$ 、边集 $\mathfrak{I} \in V \times V$ 和邻接矩阵 A 。邻接矩阵可以表示为 $A = [a_{ij}]$, 如果从节点 j 到节点 i 存在一条有向边, 则 $a_{ij} = 1$; 否则 $a_{ij} = 0$ 。有向图的拉普拉斯矩阵被定义为 $L = \text{diag}(\sum_{i=1}^N a_{ij}) - A$ 。如果存在从节点(称之为根节点)到其他节点的有向路径, 则有向图具有一个生成树。有向图 $G(V, \mathfrak{I}, A)$ 代表追随者之间的通信关系。对角增益矩阵为 $B = \text{diag}(g_i)$, 如果从领航者到追随者 i 存在有向边, 则 $g_i = 1$; 否则 $g_i = 0$ 。

下面给出常用的假设以便之后的研究。

假设 1 假定 $G(V, \mathfrak{I}, A)$ 具有一个生成树, 且在领航者节点和根节点之间至少有一个连接。

追随者的系统动态为:

$$\begin{cases} \dot{\mathbf{x}}_i = \mathbf{A}_i \mathbf{x}_i + \mathbf{B}_i \mathbf{u}_i \\ \mathbf{y}_i = \mathbf{C}_i \mathbf{x}_i, i = 1, 2, \dots, N \end{cases} \quad (1)$$

其中, $\mathbf{x}_i \in \mathbb{R}^{n_i}$ 是追随者状态, $\mathbf{u}_i \in \mathbb{R}^{m_i}$ 是追随者的控制输入, $\mathbf{y}_i \in \mathbb{R}^q$ 是追随者的输出。 $\mathbf{A}_i$ 、 $\mathbf{B}_i$ 、 $\mathbf{C}_i$ 是追随者的系统动态矩阵。

领航者的系统动态为:

$$\begin{cases} \dot{\mathbf{x}}_0 = \mathbf{S} \mathbf{x}_0 \\ \mathbf{y}_0 = \mathbf{F} \mathbf{x}_0 \end{cases} \quad (2)$$

其中, $\mathbf{x}_0 \in \mathbb{R}^n$ 、 $\mathbf{y}_0 \in \mathbb{R}^q$ 分别为领航者的状态和输出。 $\mathbf{S}$ 、 $\mathbf{F}$ 为领航者的系统动态矩阵。假设 $\mathbf{S}$ 是未知的。

假设 2 假定 $(\mathbf{A}_i, \mathbf{B}_i)$ 是可镇定的, $(\mathbf{A}_i, \mathbf{C}_i)$ 是可观测的。假设 $\mathbf{S}$ 的所有特征值都在虚轴上。

目标是设计控制器 $\mathbf{u}_i$, 使得 $\lim_{t \rightarrow \infty} \|\mathbf{y}_i(t) - \mathbf{y}_0(t)\| = 0$ 成立。

3 基于 MFADP 最优控制器的设计

3.1 观测器设计

领航者的状态以及系统动态矩阵 $\mathbf{S}$ 并不能被每一个追随者获取, 因此为每一个追随者设计观测器去预估领航者的状态以及系统动态矩阵就显得很有必要。观测器如下:

$$\dot{\mathbf{r}}_i = \mathbf{S}_i \mathbf{r}_i + \mathbf{c} \mathbf{K} [\sum_{j=1}^N (a_{ij} \mathbf{r}_j - \mathbf{r}_i) + g_i (\mathbf{x}_0 - \mathbf{r}_i)] \quad (3)$$

其中, r_i 为观测器的状态。 c 、 K 是待确定的参数, S_i 是系统动态矩阵 S 的自适应估计。令 $\rho > 0$, 自适应系统矩阵估计器被表示为:

$$\dot{S}_i = \rho \left[\sum_{j=1}^N (a_{ij} S_j - S_i) + g_i (S - S_i) \right] \quad (4)$$

在给出观测器参数设计之前, 下面的引理给出了系统动态矩阵的自适应估计结果。

引理 1^[9] 在假设 1、假设 2 成立的情况下, 给出自适应系统矩阵估计器 (式 (4)), 对于所有的 i , 可以得到系统动态矩阵误差指数收敛为 0, 即 $\lim_{t \rightarrow \infty} (S_i - S) = \lim_{t \rightarrow \infty} \tilde{S}_i = 0$ 。

定理 1 给出领航者系统 (式 (2))、观测器 (式 (3))、自适应系统矩阵估计器 (式 (4)), 设计参数 $c > 1/2\lambda_{\min}(L+B)$ 以及 $K=W$, 其中 W 是对称正定的, 且满足黎卡提方程 $WS+S^TW-W^TW+I=0$, 则观测误差以指数趋向于 0。

证明 令 $\tilde{S} = \text{col}(\tilde{S}_1, \dots, \tilde{S}_N)$ 、 $H=L+B$, 式 (4) 可以写为 $\dot{\tilde{S}} = -\rho(H \otimes I)\tilde{S}$, 在假设 1 下, H 所有的特征值都有正实部。因此, $\lim_{t \rightarrow \infty} \tilde{S}_i = 0$, 这说明对于所有的 $t \geq 0$, $\|\tilde{S}_i(t)\| \leq \beta_H \|\tilde{S}_i(0)\| e^{-\rho \alpha_H t}$, $\alpha_H > 0$, $\beta_H > 0$ 。通过式 (3), 定义观测误差为 $\tilde{r}_i = r_i - x_0$, 有:

$$\dot{\tilde{r}}_i = S\tilde{r}_i + cK \left[\sum_{j=1}^N (a_{ij} r_j - r_i) + g_i (x_0 - r_i) \right] + \tilde{S}_i r_i \quad (5)$$

令 $\tilde{r} = \text{col}(\tilde{r}_1, \dots, \tilde{r}_N)$, $\tilde{S}_d = \text{diag}(\tilde{S}_1, \dots, \tilde{S}_N)$, 然后式 (5) 可以写为下面的全局格式:

$$\dot{\tilde{r}} = ((I_N \otimes S) - c(H \otimes K))\tilde{r} + \tilde{S}_d \tilde{r} + \tilde{S}_d (1_N \otimes x_0) \quad (6)$$

其中, 1_N 是 N 维的列向量, 元素均为 1。

在假设 2 下, S 的所有特征值都在虚轴上, 当 $c > 1/2\lambda_{\min}(L+B)$ 以及 $K=W$ 时, 矩阵 $(I_N \otimes S) - c(H \otimes K)$ 是赫尔维茨矩阵, 因此根据参考文献[9], 有 $\lim_{t \rightarrow \infty} \tilde{r}_i(t) = 0$ 。

3.2 最优控制器设计

定义增广系统的状态为 $X_i = [x_i^\top, r_i^\top]^\top$, 增广系统的动态可以表示为:

$$\dot{X}_i = \begin{bmatrix} A_i & 0 \\ 0 & S \end{bmatrix} X_i + \begin{bmatrix} B_i \\ 0 \end{bmatrix} u_i + \begin{bmatrix} 0 \\ I_n \end{bmatrix} \phi = \Phi_i X_i + \bar{B}_i u_i + \bar{E} \phi \quad (7)$$

$$Y_i = C_i x_i - F r_i \quad (8)$$

其中, $\phi = cK \left[\sum_{j=1}^N (a_{ij} r_j - r_i) + g_i (x_0 - r_i) \right] + \tilde{S}_i r_i$,

$$\Phi_i = \begin{bmatrix} A_i & 0 \\ 0 & S \end{bmatrix}, \bar{B}_i = \begin{bmatrix} B_i \\ 0 \end{bmatrix}, \bar{E} = \begin{bmatrix} 0 \\ I_n \end{bmatrix}, \bar{C}_i = [C_i - F]。$$

定义折扣性能指标函数为:

$$V(X_i(t)) = \int_t^\infty e^{-\alpha_i(\tau-t)} [X_i^\top \bar{C}_i^\top Q_i \bar{C}_i X_i + u_i^\top R_i u_i] d\tau \quad (9)$$

通过对折扣性能指标函数求导获得贝尔曼方程:

$$\frac{dV_i}{dt} = \alpha_i V_i - (Y_i^\top Q_i Y_i + u_i^\top R_i u_i) \quad (10)$$

线性异构 MAS 的性能指标函数为二次的, 其表达式为:

$$V_i(X_i(t)) = X_i^\top P_i X_i \quad (11)$$

其中, P_i 为对称正定矩阵。有:

$$\begin{aligned} H(X_i, u_i) &= X_i^\top P_i (\Phi_i X_i + \bar{B}_i u_i + \bar{E} \phi) + \\ &(\Phi_i X_i + \bar{B}_i u_i + \bar{E} \phi)^\top P_i X_i - \alpha_i V_i + \\ &(Y_i^\top Q_i Y_i + u_i^\top R_i u_i) = 0 \end{aligned} \quad (12)$$

基于最优性原理, 执行 $\partial H(X_i, u_i) / \partial u_i$, 最优控制器被设计为:

$$u_i^* = -R_i^{-1} \bar{B}_i^\top P_i^* X_i \quad (13)$$

其中, P_i^* 是黎卡提方程的唯一正定解。黎卡提方程如下:

$$\Phi_i^\top P_i + P_i \Phi_i - P_i \bar{B}_i R_i^{-1} \bar{B}_i^\top P_i + \bar{C}_i^\top Q_i \bar{C}_i - \alpha_i P_i = 0 \quad (14)$$

3.3 MFADP 算法

MFADP 的基本原理将会通过下面的计算式给出。增广系统的动态 (式 (7)) 可表示为:

$$\dot{X}_i = \Phi_i X_i + \bar{B}_i (u_i - u_i^\kappa) + \bar{B}_i u_i^\kappa + \bar{E} \phi \quad (15)$$

其中, u_i^κ 是需要被更新的控制策略, u_i 是可容许的控制策略。

同时, 控制策略被更新为:

$$u_i^{\kappa+1} = -R_i^{-1} \bar{B}_i^\top P_i^\kappa X_i \quad (16)$$

于是, 获得如下计算式:

$$\begin{aligned} \frac{dV_i^\kappa}{dt} &= 2X_i^\top P_i^\kappa (\Phi_i X_i + \bar{B}_i u_i^\kappa + \bar{E} \phi) + \\ &2X_i^\top P_i^\kappa \bar{B}_i (u_i - u_i^\kappa) = \alpha_i V_i^\kappa - \end{aligned} \quad (17)$$

$$X_i^\top \bar{C}_i^\top Q_i \bar{C}_i X_i - (u_i^\kappa)^\top R_i u_i^\kappa - 2(u_i^{\kappa+1})^\top R_i (u_i - u_i^\kappa)$$

以及

$$\begin{aligned} e^{\alpha_i \Delta t} V_i^\kappa(X_i(t+\Delta t)) - V_i^\kappa(X_i(t)) = \\ \int_t^{t+\Delta t} e^{-\alpha_i(\tau-t)} (-X_i^\top \bar{C}_i^\top Q_i \bar{C}_i X_i - (u_i^\kappa)^\top R_i u_i^\kappa) d\tau + \\ \int_t^{t+\Delta t} e^{-\alpha_i(\tau-t)} (-2(u_i^{\kappa+1})^\top R_i (u_i - u_i^\kappa) + 2X_i^\top P_i^\kappa \bar{E} \phi) d\tau \end{aligned} \quad (18)$$

以上算法的实现需要借助两个神经网络, 一个神经网络用于逼近控制器, 另一个神经网络用于逼近折扣性能指标函数。然后通过式 (18), 使用最小二乘法同时获得最优控制以及最优折扣性能指标函数。

4 仿真分析

将上文的理论与算法应用到双质量-弹簧系统。用双质量-弹簧系统模型验证算法的有效性。MAS 的网络拓扑如图 1 所示。

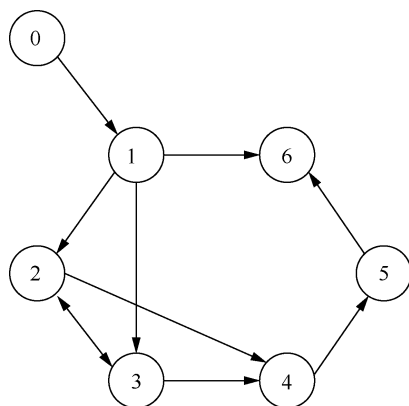


图1 MAS 的网络拓扑

追随者与领航者系统的动态矩阵分别为:

$$A_i = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{-k_i^1 - k_i^2}{m_i^1} & 0 & \frac{k_i^2}{m_i^1} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{k_i^2}{m_i^2} & 0 & -\frac{k_i^2}{m_i^2} & 0 \end{bmatrix}, \quad (19)$$

$$B_i = \begin{bmatrix} 0 \\ \frac{1}{m_i^1} \\ 0 \\ 0 \end{bmatrix}, C_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

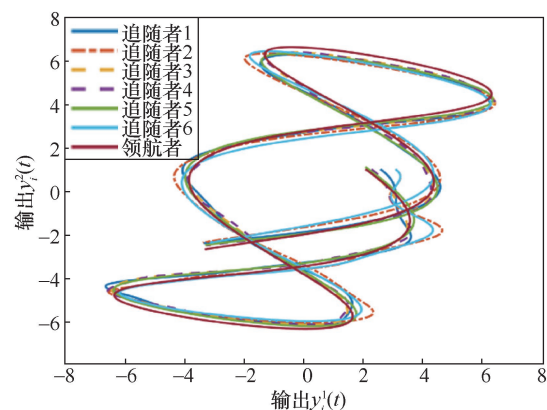
$$S = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -0.5 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 1 \\ 0.1 & 0 & -0.1 & 0 \end{bmatrix}, F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (20)$$

其中, m_i^1 和 m_i^2 是质量, 单位为 kg ; k_i^1 和 k_i^2 是两个弹簧常数。双质量-弹簧系统模型参数见表 1。

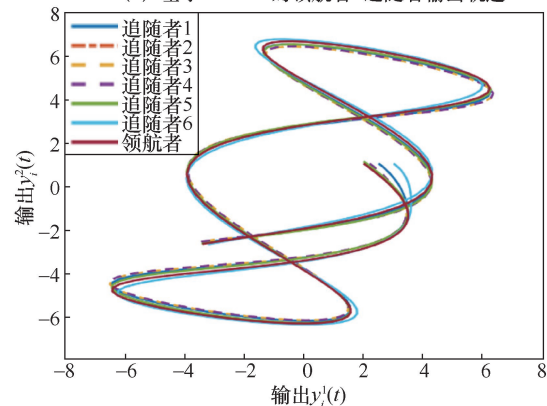
表1 双质量-弹簧系统模型参数

智能体 i	$k_i^1/(N/m)$	$k_i^2/(N/m)$	m_i^1/kg	m_i^2/kg
1	1.5	1.05	5.1	10.1
2	1.48	0.9	5.2	9.9
3	1.46	0.95	4.9	10.1
4	1.61	1.1	5.01	10.12
5	1.63	1.12	5.04	10.12
6	1.55	0.9	5.2	10.05

基于 MFADP 和 MBADP 的领航者-追随者输出轨迹如图 2 所示。从图 2 (a) 可以看出在系统动态信息未知的情况下, 追随者的输出依然可以跟踪上领航者的输出。这个实例很好地说明了该算法的有效性。



(a) 基于MFADP的领航者-追随者输出轨迹



(b) 基于MBADP的领航者-追随者输出轨迹

图2 基于 MFADP 和 MBADP 的领航者-追随者输出轨迹

5 结束语

本文针对未知异构 MAS 领航者-追随者的输出追踪问题, 运用贝尔曼最优性原理求出最优控制器, 同时还得到了代数黎卡提方程。之后基于

MFADP 算法同时求出最优控制与最优性能指标函数,值得注意的是,该算法是不需要知道系统动态的。最后通过双质量-弹簧系统模型验证了该算法的有效性。

参考文献:

- [1] LI Z K, REN W, LIU X D, et al. Distributed consensus of linear multi-agent systems with adaptive dynamic protocols[J]. Automatica, 2013, 49(7): 1986-1995.
- [2] 李金娜, 程薇燃. 基于强化学习的数据驱动多智能体系统最优一致性综述[J]. 智能科学与技术学报, 2020, 2(4): 327-340.
LI J N, CHENG W R. An overview of optimal consensus for data driven multi-agent system based on reinforcement learning[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(4): 327-340.
- [3] YANG Y L, MODARES H, WUNSCH D C, et al. Leader-follower output synchronization of linear heterogeneous systems with active leader using reinforcement learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(6): 2139-2153.
- [4] LI Q, XIA L N, SONG R Z, et al. Leader-follower bipartite output synchronization on signed digraphs under adversarial factors via data-based reinforcement learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(10): 4185-4195.
- [5] MODARES H, LEWIS F L, JIANG Z P. H_∞ tracking control of completely unknown continuous-time systems via off-policy reinforcement learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(10): 2550-2562.
- [6] 刘莹莹, 王占山. 异构多智能体系统的输出同步: 一个基于数据的强化学习方法[J]. 智能科学与技术学报, 2020, 2(4): 394-400.
LIU Y Y, WANG Z S. Output synchronization of heterogeneous multi-agent system: a reinforcement learning approach based on data[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(4): 394-400.
- [7] MODARES H, LEWIS F L, KANG W, et al. Optimal synchronization of heterogeneous nonlinear systems with unknown dynamics[J]. IEEE Transactions on Automatic Control, 2018, 63(1): 117-131.
- [8] XIA L N, LI Q, SONG R Z, et al. Optimal synchronization control of heterogeneous asymmetric input-constrained unknown nonlinear MASs via reinforcement learning[J]. IEEE/CAA Journal of Automatica Sinica, 2021, accepted.
- [9] CAI H, LEWIS F L, HU G Q, et al. The adaptive distributed observer approach to the cooperative output regulation of linear multi-agent systems[J]. Automatica, 2017, 75: 299-305.

[作者简介]



夏丽娜(1993-),女,北京科技大学自动化学院博士生,主要研究方向为多智能体同步、事件驱动控制、自适应动态规划。



李擎(1971-),男,博士,北京科技大学自动化学院教授、博士生导师,主要研究方向为群智能优化算法及其应用、多智能体同步。



宋睿卓(1982-),女,博士,北京科技大学自动化学院教授、博士生导师。主要从事基于数据的智能计算、复杂系统最优决策与控制等方面的研究工作,并将成果应用于人体信号测控、目标定位与识别等领域。共发表论文85篇,其中SCI论文44篇。第一作者ESI高被引论文3篇。出版学术专著3部,其中英文专著2部。共主持科研项目20项,包括国家自然科学基金项目、北京市自然科学基金项目、横向课题等。发表论文在Google Scholar数据库中总被引1386次,H指数19。获得2017年中国自动化学会自然科学奖一等奖。



王子涵(2000-),男,北京科技大学自动化学院在读,主要研究方向为自适应动态规划。



许镇(1986-),北京科技大学城镇化与城市安全研究院教授、博士生导师,主要研究方向为城市综合数字防灾的交叉。