DOI: 10.11992/tis.201807010
网络出版地址: http://kns.cnki.net/kcms/detail/23.1538.TP.20181230.0904.002.html

事件驱动的强化学习多智能体编队控制

徐鹏[1], 谢广明[1,2,3], 文家燕[1,2], 高远[1]
（1. 广西科技大学 电气与信息工程学院, 广西 柳州 545006; 2. 北京大学 工学院, 北京 100871; 3. 北京大学 海洋研究院, 北京 100871）

摘　要: 针对经典强化学习的多智能体编队存在通信和计算资源消耗大的问题, 本文引入事件驱动控制机制, 智能体的动作决策无须按固定周期进行, 而依赖于事件驱动条件更新智能体动作。在设计事件驱动条件时, 不仅考虑智能体的累积奖赏值, 还引入智能体与邻居奖赏值的偏差, 智能体间通过交互来寻求最优联合策略实现编队。数值仿真结果表明, 基于事件驱动的强化学习多智能体编队控制算法, 在保证系统性能的情况下, 能有效降低多智能体的动作决策频率和资源消耗。

关键词: 强化学习; 多智能体; 事件驱动; 编队控制; 马尔可夫过程; 集群智能; 动作决策; 粒子群算法
中图分类号: TP391.8　　**文献标志码:** A　　**文章编号:** 1673−4785(2019)01−0093−06

中文引用格式: 徐鹏, 谢广明, 文家燕, 等. 事件驱动的强化学习多智能体编队控制[J]. 智能系统学报, 2019, 14(1): 93–98.
英文引用格式: XU Peng, XIE Guangming, WEN Jiayan, et al. Event-triggered reinforcement learning formation control for multi-agent[J]. CAAI transactions on intelligent systems, 2019, 14(1): 93–98.

Event-triggered reinforcement learning formation control for multi-agent

XU Peng[1], XIE Guangming[1,2,3], WEN Jiayan[1,2], GAO Yuan[1]
(1. School of Electric and Information Engineering, Guangxi University of Science and Technology, Liuzhou 545006, China; 2. College of Engineering, Peking University, Beijing 100871, China; 3. Institute of Ocean Research, Peking University, Beijing 100871, China)

Abstract: A large consumption of communication and computing capabilities has been reported in classical reinforcement learning of multi-agent formation. This paper introduces an event-triggered mechanism so that the multi-agent's decisions do not need to be carried out periodically; instead, the multi-agent's actions are replaced depending on the event-triggered condition. Both the sum of total reward and variance in current rewards are considered when designing an event-triggered condition, so a joint optimization strategy is obtained by exchanging information among multiple agents. Numerical simulation results demonstrate that the multi-agent formation control algorithm can effectively reduce the frequency of a multi-agent's action decisions and consumption of resources while ensuring system performance.

Keywords: reinforcement learning; multi-agent; event-triggered; formation control; Markov decision processes; swarm intelligence; action-decisions; particle swarm optimization

强化学习是受动物能有效适应环境的启发发展而来的一种算法。基本思想是以试错的机制与环境进行交互, 在没有导师信号的情况下, 使奖励累积最大化, 来寻求最优的策略[1-3]。目前强化学习的行业应用颇广泛, 比如无人驾驶、人形机器人、智能交通和多智能体协同等。其中多智能体编队的强化学习研究是一个重要的方向[4-5]。文献[4]设计多动作回放的马尔可夫模型, 在此框架下, 多智能体 Q 学习可收敛到最优的联合行动策略。文献[5]提出一种评估 Q 值法, 多智能体通

收稿日期: 2018−07−11.　**网络出版日期:** 2019−01−03.
基金项目: 国家重点研发计划项目 (2017YFB1400800); 国家自然科学基金项目 (91648120, 61633002, 51575005, 61563006, 61563005); 广西高校工业过程智能控制技术重点实验室项目 (IPICT-2016-04).
通信作者: 文家燕. E-mail: wenjiayan2012@126.com.

过交流 Q 值函数和折扣奖励方差来学习世界，较快完成了编队任务。然后，智能体在这些学习过程中，需要连续地与环境进行交互，会导致大量的通信和计算资源消耗。因此在有限资源情况下，保证多智能体系统的编队性能，考虑如何降低资源消耗是必要的，这也是促使开展本项研究的直接原因。

事件驱动机制已经被证明可以有效地减小大规模网络的通信量[6-7]。综合已有研究成果，事件驱动条件设计主要分为两类：状态相关[8]和状态无关[9]。其主要做法都是通过检测智能体采样前后状态的偏差值大小，判断是否满足事件驱动条件，来决定间歇性的更新控制输入，减小控制器与多智能体系统的通信频率和计算量[10-12]。文献[10]较早在状态反馈控制器中引入事件驱动控制机制。文献[11]考虑多智能体间同步采样异步触发机制解决多智能体环形编队问题，其中智能体可独立地选择触发条件参数。但是当前强化学习与事件驱动的结合相对较少[13-14]。文献[13]设计事件驱动控制器并应用于非线性连续系统的强化学习中，解决了自适应动态规划问题。文献[14]提出根据智能体观测信息的变化率设计触发函数，减少学习过程中的计算资源消耗。

综合以上分析，本文区别于传统的多智能体强化学习算法，在资源有限的情况下，考虑将事件驱动和强化学习相结合，侧重于事件驱动在强化学习过程中动作决策频率方面的研究。

1　问题描述

1.1　基于强化学习的编队问题

$Z_{>0}$ 表示正整数集合。多智能体编队问题描述如图 1 所示，假设有 $N(N\geqslant 2)$ 个智能体从初始位置出发，初始位置为随机分布的坐标点，抵达各自的期望位置，每个智能体对应的期望位置点不同，且期望位置点的数量等于多智能体个体的数量。为便于分析，令多智能体在二维网格中运动，定义网格的大小为 $(x_{t,i},y_{t,i})$，在二维网格中坐标 $(x_{t,i},y_{t,i})$ 表示智能体 i 的状态 $s_{t,i}\in Z_{>0}^2$，并朝着对应的期望点 $G_i(i=1,2,\cdots,N)$ 运动，每个智能体对应的期望位置点会按照贪婪法则预先给定。智能体 i 运动过程中，动作集合 $A_i(s)$={上，下，左，右，保持}，因此下一刻状态值可能为 $(x_{t,i},y_{t,i}-1)$，$(x_{t,i},y_{t,i}+1)$，$(x_{t,i}-1,y_{t,i})$，$(x_{t,i}+1,y_{t,i})$ 和 $(x_{t,i},y_{t,i})$。

综上所述，基于强化学习的多智能体编队问题可描述为：智能体与环境进行交互，学习动作决策策略，最小化群体的动作总量 K_1，使多智能体抵达各自的期望位置点，且在运动过程中不发生碰撞。在学习最优动作策略过程中，当所有智能体都抵达目标位置时，群体会得到一个 r^+ 的奖励，否则会得到 r^- 的惩罚。每个智能体都可以与邻居智能体交互，来获取其他智能体的奖励信号。

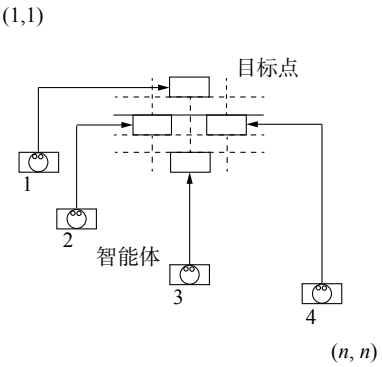


图 1　编队问题

Fig. 1　Formation problem

1.2　分散式马尔可夫模型

博弈论中智能体的每一个决策都会导致状态的转移，此时的决策序列称为一个随机策略。具有马尔可夫 (Markov) 特性的随机策略称为 Markov 策略 (MG)。MG 是研究具有离散时间特性的多智能体系统的重要理论框架。

考虑一个分散式马尔可夫模型 (decentralized Markov decision processes, DEC-MDPs)，DEC-MDPs 是一个五元组<I, S, $A_i(s)$, P, $\{r_i\}$>，其中：I 为有限智能体集合；S 为状态集合；$A_i(s)$ 为第 i 个智能体在状态 $s\in S$ 下可选动作集合，则多智能体在状态 s 下的联合行动表示为 $A(s)=A_1(s)\times A_2(s)\times\cdots\times A_N(s)$；$P$ 为动作转移概率；r_i 表示智能体 i 奖赏值。在 DEC-MDPs 中，每个智能体不依赖全局信息，只保持自身和编队期望点的相对关系，且每个智能体只需获取自身的局部观测信息，在通信无障碍情况下，这些局部信息的并集为一个完整的全局信息。多智能体的混合策略组合 $\pi=(\pi_1(s),\pi_2(s),\cdots,\pi_N(s))$ 构成整个系统的一个混合策略，策略 π 可看成是状态空间 s 到动作空间 $A_i(s)$ 的映射。求解 DEC-MDPs 的目的是寻求一个最优策略 π，来最大化系统的回报值。

1.3　Q 学习

Q 学习是最早的在线强化学习算法，同时也是强化学习最重要的算法之一。Watkins 在博士论文中[15]提出了 Q 学习算法，如图 2 所示，通过与环境的交互，学习环境状态到行为的映射关系，使智能体从环境中获得最大累积奖赏值，通常用值函数来评价策略 π 的优劣。

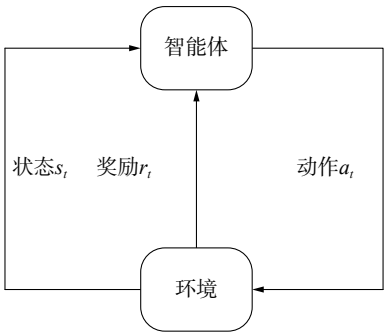


图 2　Q 学习流程图

Fig. 2　Flow chart of Q-learning

图 2 采用的是折扣累积奖赏, 策略 π 的状态值函数为

$$V^{\pi}(s)=\sum_{t=0}^{\infty}\gamma^{t}r_t(s_t,a_t)|s=s_0,\quad a_t=\pi(s_t)$$

式中, γ 为折扣因子, s_0 为初始状态。另一种形式的值函数是状态动作值函数:

$$Q^{\pi}(s_t,a_t)=r(s_t,a_t)+\gamma V^{\pi}(s_{t+1})$$

此时最优策略可以根据式 (1) 得到:

$$\pi^{*}=\arg\max_{a\in A(s)}Q^{\pi}(s,a)\tag{1}$$

那么可借助时间差分误差来更新 Q 函数, 智能体将观测到的数据代入 Q 函数中进行迭代学习, 得到精确的解:

$$\nabla Q_{t+1}(s_t,a_t)=r_{t+1}+\gamma\max Q_t(s_{t+1},a')-Q_t(s_t,a_t)$$

$$Q_{t+1}(s_t,a_t)=Q_t(s_t,a_t)+\alpha_t\nabla Q_{t+1}(s_t,a_t)$$

式中, t 为当前时刻; α_t 为当前的学习率; a' 为状态 $s+1$ 时执行的动作。

Q 学习最大的特点是智能体可以通过试错的方式寻求最优的策略, 因此所有的状态动作都需要被无限次地遍历, 同时这也会造成大量的通信和计算资源消耗。

2　算法设计

为解决经典强化学习过程中存在通信和计算资源消耗大问题, 本节在经典强化学习中引入事件驱动控制机制。

2.1　事件驱动条件设计

在 DEC-MDPs 中, 每个智能体可独立地观测局部状态信息, 同时广播给附近的其他智能体。观测结束后, 其根据上一时刻观测与当前观测的状态偏差值大小, 决定是否要执行更新动作。这里采用状态值 $Q_{t,i}(s_{t,i},a_{t,i})$ 作为智能体 i 在 t 时刻的当前观测值, $Q_{t,i}(s_{t,i},a_{t,i})$ 可通过查询 Q-Table 获得, 则智能体 i 从 $t-1$ 时刻到 t 时刻的偏差值可写成:

$$e_i(t)=Q_{t,i}(s_{t,i},a_{t,i})-Q_{t-1,i}(s_{t-1,i},a_{t-1,i})$$

式中, $t>0$; $e_i(t)$为观测量的状态偏差值; $Q_{t-1,i}(s_{t-1,i},a_{t-1,i})$为 $t-1$时刻状态观测值。

在基于事件驱动的强化学习编队问题中, 如果智能体 i 在期望位置点上, 会获得较大的奖赏值。换句话说, 当智能体 i 迅速到达期望位置时, 获得累积折扣奖赏值较大。因此, 根据智能体的累积折扣奖赏值进行设计状态阈值函数是合理的。但是, 状态阈值函数如果仅通过累积折扣奖赏值去评估, 智能体 i 往往会获得自私的策略, 不利于学到群体最优的策略。因此, 考虑在智能体 i 的状态阈值函数中引入当前奖励的偏差 $\delta_{t,i}$。假设智能体能观测到周围的一圈 10 个格子, 如果智能体 j 存在于智能体 i 的观测范围内, 称智能体 j 为智能体 i 的邻居, 则 t 时刻智能体 i 奖励的偏差 $\delta_{t,i}$ 可写成:

$$\delta_{t,i}=\left(\frac{\sum\limits_{j\in N_{t,i}}r_{t,j}-|N_{t,i}|r_{t,i}}{|N_{t,i}|}\right)^{2}$$

式中, $N_{t,i}$ 为智能体 i 在 t 时刻邻居集合, $|N_{t,i}|$为智能体 i 邻居个数。当智能体 i 的状态偏差大于状态阈值函数时, 更新智能体 i 的动作并对自身动作决策进行广播。同一时刻里, 不一定所有的智能体都会被驱动, 未被驱动的智能体仅接受信息, 有利于减少多智能体系统通信和计算资源的消耗, 则事件驱动条件设计为式 (2):

$$e_i(t)=\sigma_i\left(Q_{t,i}(s_{t,i},a_{t,i})-\sqrt{\delta_{t,i}}\right)\tag{2}$$

式中 $0<\sigma_i<1$。

2.2　基于事件驱动的 Q 学习

如图 3 所示, 智能体执行过程为: 当智能体感知自己附近有障碍物时, 即优先避碰, 避碰结束后重新进行编队。在通信无障碍的情况下, 考虑基于事件驱动的 DEC-MDPs, 由六元组<I, S, $A_i(s)$, P, $\{r_i\}$, e>构成, 其中 e 表示状态偏差值, 当 $e_i(t)>\sigma_i\left(Q_{t,i}(s_{t,i},a_{t,i})-\sqrt{\delta_{t,i}}\right)$ 时, 智能体 i 更新动作, 否则执行上一刻的采样动作。

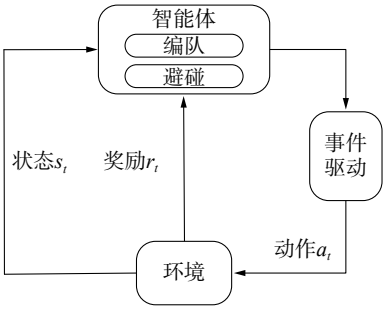


图 3　基于事件驱动的强化学习框架

Fig. 3　The frame of reinforcement learning with event-triggered

图 2 经典的 Q 学习是使用一个合理的策略产生动作, 根据动作与环境交互, 可得到下一刻的

状态以及奖赏值，不断地优化奖赏值来得到最优的 Q 函数。基于事件驱动的 Q 学习不同于经典的 Q 学习算法，智能体首先判断事件条件是否触发，来决定是否基于当前的状态值，更新动作与环境进行交互。多智能体从各自的初始位置点出发，当每个智能体都抵达期望位置点时，称为一轮 Episode 学习终止。则对于事件驱动的强化学习多智能体编队算法可描述为

1) 初始化 Q 矩阵；

2) 初始化多智能体的当前状态 s_0；

3) 智能体 i 以行为策略 ($\varepsilon-$ 贪心策略) 选择动作 $a_{t,i}$；

4) 智能体 i 与环境交互，获取下一个状态 s'_t 和即时的奖赏值 $r_{t,i}$；

5) 智能体 i 更新当前状态 $s_{t,i}$ 和 $a_{t,i}$ 的 $Q_{t,i}(s_{t,i},a_{t,i})$ 值：

$$\Delta Q_{t,i}(s_{t,i},a_{t,i}) = r_{t,i}+\gamma\max Q_{t,i}(s'_{t,i},a_{t,i})-Q_{t,i}(s_{t,i},a_{t,i})$$

$$Q_{t,i}(s_{t,i},a_{t,i}) \leftarrow Q_{t,i}(s_{t,i},a_{t,i})+\alpha\Delta Q_{t,i}(s_{t,i},a_{t,i})$$

式中，$0<\alpha<1$ 为学习率，$0<\gamma<1$ 为折扣因子；

6) 如果每个智能体都抵达各自期望位置，则终止一轮 Episode；

7) 判断是否满足事件触发条件，如果满足返回步骤 3)，不满足返回步骤 4)。

2.3　资源消耗对比

Q 学习中计算资源消耗，主要体现在遍历所有的策略来寻求最优解。每次学习过程中，智能体都要基于当前的状态遍历 $Q(s,a)$ 值表，查找一个最优的策略。$Q(s,a)$ 值表的实现采用 Lookup 表格，其中 $s\in S$ 和 $a\in A_i(s)$，表的大小为 $S\times A$ 的乘积的元素个数。下面举例说明 $Q(s,a)$ 表大小，假设存在 N 个智能体，每个智能体有 M 个动作，环境中共存在 n^2 状态，那么 $Q(s,a)$ 值表的大小为 $M^N\times n^{2N}$，在 ρ 步中，智能体共需遍历 $M^N\times n^{2N}\times\rho$ 次 $Q(s,a)$ 值表，做 $M\rho$ 次动作决策，这需要占用极大的通信和计算资源。假设智能体在在 ρ 步中，有 λ 次不被驱动，则通信次数减少为 $M(\rho-\lambda)$ 次，遍历次数减少为 $M^N\times n^{2N}\times(\rho-\lambda)$ 次。虽然基于事件驱动的强化学习压缩了整个学习的解空间，但在计算和通信资源限制下，基于事件驱动的强化学习通过减少智能体的动作决策，能在短时间内找到一个动作总量为 K_2 的可行的编队策略，通过不断更新迭代，最终寻求到最小化群体的动作总量。

3　数值仿真分析

为了定量比较经典 Q 学习和事件驱动 Q 学习动作决策频率的大小，假设智能体随机初始化在大小为 20×20 的格子世界中，如图 4 所示，存在 3 个智能体，每个智能体动作集合为 $A_i(s)$，可观测到格子的可能为“障碍物”“目标点”“普通格子”，奖赏 r_i 可写成式 (3):

$$r_i=\omega+0.1\beta+\chi+\xi \tag{3}$$

式中，当智能体 i 未抵达目标点时 ω 为−1，否则为 0；智能体 i 抵达期望点时 β 为−1，否则为 0；当智能体 i 移动到边界或者撞上 L 型的障碍物时，智能体保持不动，此时 χ 为−1，否则为 0；当智能体 i 与智能 j 在同一时刻向同一格子移动时，两个智能体保持不动，此时 ζ 为−1，否则为 0。

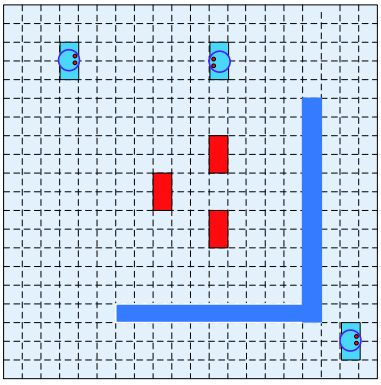


图 4　多智能体编队

Fig. 4　Formation of multi-agents systems

表 1 比较了事件驱动 Q 学习和经典 Q 学习的动作决策次数，为方便表达做如下定义：κ_1 为经典 Q 学习决策次数，κ_2 为事件驱动 Q 学习决策次数，则减少决策率 η 可由如式 (4) 计算:

表 1　事件驱动与经典 Q 学习动作决策次数对比

Table 1　A comparison of action times between event-triggered Q and classical Q case

经典 Q 学习	$\sigma_i=0.05$		$\sigma_i=0.02$		$\sigma_i=0.01$	
	事件驱动 Q 学习	减少决策率 η /%	事件驱动 Q 学习	减少决策率 η /%	事件驱动 Q 学习	减少决策率 η/%
600 000	192 976	67.79	129 628	78.32	100 143	83.31
900 000	372 463	58.61	257 586	71.37	246 968	72.56
1 200 000	591 433	50.71	475 579	60.36	406 864	66.10
1 500 000	828 361	44.77	689 172	54.05	587 553	60.83

$$\eta = \frac{K_1 - K_2}{K_1} \tag{4}$$

在同一组 $\sigma_i = 0.05$ 参数下, 随着 Episode 的增加, 事件驱动 Q 学习减少的决策次数从 407 024 次增加到 671 639 次, 但减少决策率 η 却从 67.79% 下降到 44.77%, 可得减少决策次数的增长率逐渐下降。因此随着算法的渐近收敛, 减少的决策次数会趋近于一个饱和值。在同样 Episode 下, 可得不同 σ_i 值的事件驱动条件都能减少学习过程中的动作决策频率。

如图 5 所示, 基于事件驱动 Q 学习和经典 Q 学习经过 200 轮 Episode 训练, 可得 $K_1 \approx K_2$, 说明两种算法都成功完成编队任务, 且完成编队任务的动作次数趋近一致。相比经典 Q 学习, 基于事件驱动 Q 学习曲线梯度下降快, 说明该算法能较快地找到一个成功策略, 完成编队任务。随着减少的决策次数趋近稳定, 解空间被释放, 通过不断迭代更新, 寻求到最优解。

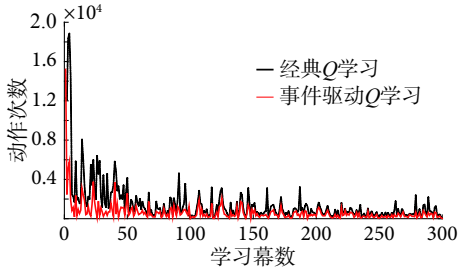


图 5　基于事件驱动 Q 学习与经典 Q 学习动作次数演变
Fig. 5　Variation of the number of actions of event-triggered Q and classical Q

图 6 对比了在不同参数下的事件驱动条件编队动作次数的演变情况。结合表 1, 在一个 Episode 中, 虽然 σ_i 参数变小会降低编队系统的决策率, 但同时也会增加编队的动作决策次数, 在基于事件驱动的学习过程中, 当 $K_2(1-\eta) < K_1$ 事件驱动函数被视为有效。

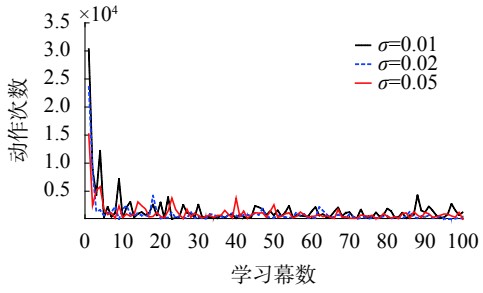


图 6　基于事件驱动 Q 学习不同 σ_i 下的动作次数变化
Fig. 6　Variation of the number σ_i of actions of event-triggered Q

4　结束语

本文主要研究基于事件驱动的强化学习多智能体编队问题, 侧重于学习过程中动作决策层面的研究。智能体在与环境交互中, 根据观测状态值的变化与设计的事件驱动条件比较, 决定是否执行动作更新。研究结果表明, 在相同时间内, 保证系统可允许编队性能的前提下, 事件驱动机制可以降低智能体的动作决策频率和减少通信和计算资源消耗。因此, 引入事件驱动机制有助于强化学习在实际有限资源环境中的工程应用。未来的工作会基于现有研究, 将事件驱动机制优势与更多种类的强化学习算法相结合, 开展相关的理论和应用研究。

参考文献:

[1] POLYDOROS A S, NALPANTIDIS L. Survey of model-based reinforcement learning: applications on robotics[J]. Journal of intelligent & robotic systems, 2017, 86(2): 153–173.

[2] TSAURO G, TOURCTZKY D S, LN T K, et al. Advances in neural information processing systems[J]. Biochemical and biophysical research communications, 1997, 159(6).

[3] 梁爽, 曹其新, 王雯珊, 等. 基于强化学习的多定位组件自动选择方法[J]. 智能系统学报, 2016, 11(2): 149–154.
LIANG Shuang, CAO Qixin, WANG Wenshan, et al. An automatic switching method for multiple location components based on reinforcement learning[J]. CAAI transactions on intelligent systems, 2016, 11(2): 149–154.

[4] KIM H E, AHN H S. Convergence of multiagent Q-learning: multi action replay process approach[C]//Proceedings of 2010 IEEE International Symposium on Intelligent Control. Yokohama, Japan, 2010: 789–794.

[5] IIMA H, KUROE Y. Swarm reinforcement learning methods improving certainty of learning for a multi-robot formation problem[C]//Proceedings of 2015 IEEE Congress on Evolutionary Computation. Sendai, Japan, 2015: 3026–3033.

[6] MENG Xiangyu, CHEN Tongwen. Optimal sampling and performance comparison of periodic and event based impulse control[J]. IEEE transactions on automatic control, 2012, 57(12): 3252–3259.

[7] DIMAROGONAS D V, FRAZZOLI E, JOHANSSON K H. Distributed event-triggered control for multi-agent systems[J]. IEEE transactions on automatic control, 2012, 57(5): 1291–1297.

[8] XIE Duosi, XU Shengyuan, CHU Yuming, et al. Event-triggered average consensus for multi-agent systems with nonlinear dynamics and switching topology[J]. Journal of the franklin institute, 2015, 352(3): 1080–1098.

[9] WU Yuanqing, MENG Xiangyu, XIE Lihua, et al. An input-based triggering approach to leader-following prob-

lems[J]. Automatica, 2017, 75: 221–228.

[10] TABUADA P. Event-triggered real-time scheduling of stabilizing control tasks[J]. IEEE transactions on automatic control, 2007, 52(9): 1680–1685.

[11] WEN Jiayan, WANG Chen, XIE Guangming. Asynchronous distributed event-triggered circle formation of multi-agent systems[J]. Neurocomputing, 2018, 295: 118–126.

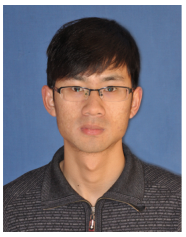
[12] MENG Xiangyu, CHEN Tongwen. Event based agreement protocols for multi-agent networks[J]. Automatica, 2013, 49(7): 2125–2132.

[13] ZHONG Xiangnan, NI Zhen, HE Haibo, et al. Event-triggered reinforcement learning approach for unknown nonlinear continuous-time system[C]//Proceedings of 2014 International Joint Conference on Neural Networks. Beijing, China, 2014: 3677–3684.

[14] 张文旭, 马磊, 王晓东. 基于事件驱动的多智能体强化学习研究[J]. 智能系统学报, 2017, 12(1): 82–87.
ZHANG Wenxu, MA Lei, WANG Xiaodong. Reinforcement learning for event-triggered multi-agent systems[J]. CAAI transactions on intelligent systems, 2017, 12(1): 82–87.

[15] KRÖSE B J A. Learning from delayed rewards[J]. Robotics and autonomous systems, 1995, 15(4): 233–235.

作者简介：



徐鹏，男，1991 年生，硕士研究生，主要研究方向为多智能体、强化学习、深度学习。

谢广明，男，1972 年生，教授，博士生导师，主要研究方向为复杂系统动力学与控制、智能仿生机器人多机器人系统与控制。现主持国家自然基金重点项目 3 项，发明专利授权 10 余项。曾荣获教育部自然科学奖一等奖、国家自然科学奖二等奖。发表学术论文 300 余篇，其中被 SCI 收录 120 余篇、EI 收录 120 余篇。

文家燕，男，1981 年生，副教授，博士，主要研究方向为事件驱动控制、多智能体编队控制。发表学术论文 10 余篇。