SCIENCE CHINA
Information Sciences

• LETTER •

August 2020, Vol. 63 189205:1–189205:3
https://doi.org/10.1007/s11432-018-9671-2

Data-driven containment control of discrete-time multi-agent systems via value iteration

Zhinan PENG[1], Jiangping HU[1*] & Bijoy Kumar GHOSH[1,2]

[1]*School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu* 611731*, China;*
[2]*Department of Mathematics and Statistics, Texas Tech University, Lubbock* 79409*, USA*

Received 6 August 2018/Revised 17 October 2018/Accepted 21 November 2018/Published online 17 April 2020

Citation Peng Z N, Hu J P, Ghosh B K. Data-driven containment control of discrete-time multi-agent systems via value iteration. Sci China Inf Sci, 2020, 63(8): 189205, https://doi.org/10.1007/s11432-018-9671-2

Dear editor,

We consider a containment control problem for a linear discrete-time (DT) multi-agent system via reinforcement learning. Containment control (CC) of multi-agent networks has received extensive attention in the control community in recent years [1–4]. CC is motivated by natural phenomena and has potential and vital applications in practical engineering. For example, one application of CC is to ensure that no member of a group of robots or autonomous vehicles enters a dangerous area. In that case, part of the group is introduced as a leader to enable the vehicles or robots to enter the safety zone. Most previous work on CC has considered continuous-time (CT) multi-agent systems, in which the aim is to drive the states or outputs of all the followers to the convex hull spanned by the states or outputs of the leaders. Recently some results were reported regarding heterogeneous followers with non-identical dynamics in [5]. Till now, it is still very challenging to design CC algorithms for multi-agent systems when the agent dynamics are completely unknown. To this end, we need to consider a data-driven design for containment control. Therefore, the aim of the present study is to design a distributed algorithm that not only solves CC problems but also minimizes a local energy cost by using only input and output data and provide a solution for containment control problems of multi-agent systems under the framework of reinforcement learning.

Data-driven containment control. Consider a discrete-time multi-agent system with N followers and M leaders. The dynamics of the followers are given as follows:

$$x_i(k+1)=Ax_i(k)+B_iu_i(k),\ i\in\mathcal{F}=\{1,\dots,N\},\ (1)$$

where $x_i(k)\in\mathbb{R}^n$ denotes the state, $u_i(k)\in\mathbb{R}^{p_i}$ is the control input, $A\in\mathbb{R}^{n\times n}$ and $B_i\in\mathbb{R}^{n\times p_i}$ are constant system matrices. The dynamics of the leaders are given as follows:

$$w_m(k+1)=Aw_m(k),\ m\in\mathcal{H}=\{1,\dots,M\},\ (2)$$

where $w_m(k)\in\mathbb{R}^n$ is the state of the m-th leader.

For the considered multi-agent system (1) and (2), our objective is to design a distributed control input $u_i(k)$ for each follower by only using information from its neighbours to minimize a performance index function $J_i(x_i(k),u_i(k))$, and simultaneously, guarantee that the states of all the followers will converge to the convex hull spanned by the states of the leaders, that is, $\forall i\in\mathcal{F}$,

$$\lim_{k\to\infty}\operatorname{dist}\big(x_i(k),Co(w_1(k),\dots,w_M(k))\big)=0.$$

We use a directed graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ to model the interaction network associated with the multi-agent system (1) and (2), where $\mathcal{V}=\{1,\dots,N,N+1,\dots,N+M\}$ is the set of nodes, and $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ is the set of arcs. Simultaneously, we use $\mathcal{G}^f=(\mathcal{V}(\mathcal{G}^f),\mathcal{E}(\mathcal{G}^f))$ to represent the network topology of N followers, and thus

* Corresponding author (email: hujp@uestc.edu.cn)

$\mathcal{V}(\mathcal{G}^f) = \mathcal{F} \subseteq \mathcal{V}$. A weighted adjacency matrix $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{N\times N}$ of the subgraph $\mathcal{G}^f$ has nonnegative elements, which is positive; i.e., $a_{ij} > 0$, if and only if $(v_j, v_i) \in \mathcal{E}(\mathcal{G}^f)$; $a_{ij} = 0$, otherwise. The set $\mathcal{N}(i) = \{v_j \in \mathcal{V} \,|\, (v_j, v_i) \in \mathcal{E}\}$ denotes the set of neighbors of agent i. A diagonal matrix D is an in-degree matrix of $\mathcal{G}^f$, whose diagonal elements $d_i = \sum_{j\in\mathcal{N}(i)} a_{ij}$ for $i = 1, 2, \ldots, N$. The Laplacian matrix $\mathcal{L}$ of the digraph $\mathcal{G}^f$ is defined by $\mathcal{L} = D - \mathcal{A} \in \mathbb{R}^{N\times N}$. Further, a diagonal matrix $\mathcal{B}_m = \mathrm{diag}\{b_1^m, \ldots, b_N^m\} \in \mathbb{R}^{N\times N}$ denotes the leader adjacency matrix associated with the m-th leader for $m = 1, \ldots, M$, where $b_{im} > 0$ if $(v_m, v_i) \in \mathcal{E}$; i.e., the m-th leader is a neighbor of agent i; and $b_{im} = 0$, otherwise.

For the i-th follower, a disagreement vector is defined as follows:

$$\varepsilon_i(k) = \sum_{j=1}^{N} a_{ij}(x_i(k) - x_j(k)) + \sum_{m=1}^{M} b_i^m(x_i(k) - w_m(k)).$$

Let $\varepsilon(k) = (\varepsilon_1^{\mathrm{T}}, \varepsilon_2^{\mathrm{T}}, \ldots, \varepsilon_N^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{nN}$ be the overall disagreement vector. Then we have the following compact form:

$$\varepsilon(k) = \sum_{m=1}^{M} \big((\mathcal{H}_m) \otimes I_n\big)(x(k) - \bar{w}_m(k)\big),$$

where $\bar{w}_m(k) = 1_N \otimes w_m(k)$, and $\mathcal{H}_m = \frac{1}{M}\mathcal{L} + \mathcal{B}_m$.

For each agent i, we define a discounted performance index function as follows:

$$J_i(\varepsilon_i(k), u_i(k)) = \sum_{n=k}^{\infty} \alpha^{n-k} c_i(\varepsilon_i(n), u_i(n), u_{\mathcal{N}(i)}(n)), \quad (3)$$

where $c_i(\varepsilon_i(k), u_i(k)) = \varepsilon_i^{\mathrm{T}}(k) Q_{ii} \varepsilon_i(k) + u_i^{\mathrm{T}}(k) R_{ii} \cdot u_i(k) + \sum_{j\in\mathcal{N}(i)} u_j^{\mathrm{T}}(k) S_{ij} u_j(k)$ is the utility function, $\mathcal{N}(i) = \{j \in \mathcal{V} \,|\, (j, i) \in \mathcal{E}\}$ denotes the set of neighbors of agent i, and $u_{\mathcal{N}(i)}(k)$ are the control inputs of the neighbors of agent i, $\alpha \in (0, 1]$ is discount factor, $Q_{ii} > 0$, $R_{ii} > 0$ and $S_{ij} > 0$ are symmetric positive definite matrices with suitable dimensions. According to the Bellman's optimality principle, the optimal index function J_i^* satisfies the coupled discrete-time Hamilton-Jacobi-Bellman (DT-HJB) equation

$$J_i^*(\varepsilon_i(k)) = \min_{u_i(k)} \{c_i(\varepsilon_i(k), u_i(k)) + \alpha J_i^*(\varepsilon_i(k+1))\}. \quad (4)$$

Then one can have the following optimal control law:

$$u_i^*(k) = -\frac{\alpha}{2}\left(d_i + \sum_{m=N+1}^{N+M} b_i^m\right) \cdot R_{ii}^{-1} B_i^{\mathrm{T}} \nabla J_i^*(\varepsilon_i(k+1)). \quad (5)$$

The optimal containment control solution (5) relies on solving the DT-HJB equation (4), but the exact solution of the equation is generally impossible to be obtained. We now present a value iteration (VI) algorithm (Algorithm 1) to compute the solution of the coupled DT-HJB equation (4). The optimal performance index function $J_i^*(\varepsilon_i(k))$ and the optimal control law $u_i^*(k)$ are respectively approximated by the iterative performance index function $J_i^l(k)$ and the iterative control law $u_i^l(k)$ with the iteration index l.

Algorithm 1 VI algorithm for the DT-HJB equation

Initialization: For $\forall i$, let $l = 0$. Start with any initial control law $u_i^0(k)$ and performance index function $J_i^0(k)$. Choose an arbitrarily computation precision ϵ.

1: **repeat**

2: The iterative performance index function $J_i^l(k)$ is updated according to the Bellman equation as follows:

$$J_i^{l+1}(k) = c_i(\varepsilon_i(k), u_i^l(k)) + \alpha J_i^l(k+1); \quad (6)$$

3: The iterative control law $u_i^l(k)$ is updated according to the following scheme:

$$u_i^{l+1}(k) = \arg\min_{u_i(k)} \{c_i(\varepsilon_i(k), u_i(k)) + \alpha J_i^{l+1}(k+1)\}; \quad (7)$$

4: **until** on convergence of $|J_i^{l+1}(k) - J_i^l(k)| \leqslant \epsilon$; end.

Theorems 1 and 2 give the convergence results of the proposed VI algorithm and the stability analysis of the closed-loop multi-agent system under the optimal containment control laws (5). The proofs of Theorems 1 and 2 are given in Appendix A.

Theorem 1. Assume there exist admissible control laws $u_i^l(k)$ for $\forall i$. When each leader has a directed path to all the followers, then one has $J_i^l(k)$ and $u_i^l(k)$ will respectively converge to the optimal solution $J_i^*(\varepsilon(k))$ and $u_i^*(k)$, as $l \to \infty$.

Theorem 2. When each leader has a directed path to all the followers, then all the states of the followers converge to the convex hull spanned by the sates of the leaders under the optimal control laws (5), as $k \to \infty$.

Actor-critic online learning. Because the iterative performance index function $J_i^l(k)$ and the iterative control law $u_i^l(k)$ depend on the information of the state matrices B_i, in order to use only the measured state information, we use an actor-critic neural network to approximate $J_i^l(k)$ and $u_i^l(k)$ in real implementation. First, a critic network (CN) is used to approximate the iterative performance index function $J_i^l(k)$ and can be expressed by

$$\hat{J}_i(k) = \bar{X}_i^{\mathrm{T}}(k) \hat{W}_{ci} X_i(k), \quad (8)$$

where $\hat{W}_{ci} \in \mathbb{R}^{n\times nq}$ is the critic weight, q is the number of agent i and its neighbors, $\bar{X}_i$ is the vector ε_i, X_i is a stack vector of ε_i and its neighbors'

states. The output of the CN is given by

$$\mathcal{T}_{J_i}(k) = \varepsilon_i^{\mathrm{T}}(k)Q_{ii}\varepsilon_i(k) + \hat{u}_i^{l\mathrm{T}}(k)R_{ii}\hat{u}_i^l(k) + \sum_{j\in\mathcal{N}(i)} \hat{u}_j^{l\mathrm{T}}(k)S_{ij}\hat{u}_j^l(k) + \hat{J}_i^l(k+1). \quad (9)$$

The approximation error is defined by $e_{ci}(k) = \mathcal{T}_{J_i}(k) - \hat{J}_i(k)$. The objective function to be minimized in the CN training is given by $E_{ci}(k) = \frac{1}{2}e_{ci}^{\mathrm{T}}(k)e_{ci}(k)$. We use the gradient descent method to update the CN weights. Second, an actor network (AN) is used to approximate the iterative control law $u_i^l(k)$ and can be expressed by

$$\hat{u}_i(k) = \hat{W}_{ai}^{\mathrm{T}}X_i(k), \quad (10)$$

where $\hat{W}_{ai} \in \mathbb{R}^{nq\times p_i}$ is the actor weight. The approximation error defined for the AN is given by $e_{ai}(k) = \hat{u}_i(k) - \overline{u}_i(k)$, where $\overline{u}_i(k)$ is the target control law of the AN according to (5), which is given by $\overline{u}_i(k) = -\frac{\alpha}{2}(d_i + \sum_{m=N+1}^{N+M} b_i^m)R_{ii}^{-1}B_i^{\mathrm{T}}\frac{\partial \hat{J}_i(\varepsilon_i(k+1))}{\partial \varepsilon_i(k+1)}$. The objective error function to be minimized in the AN training is given by $E_{ai}(k) = \frac{1}{2}e_{ai}^{\mathrm{T}}(k)e_{ai}(k)$. The implement process is shown in Appendix B.

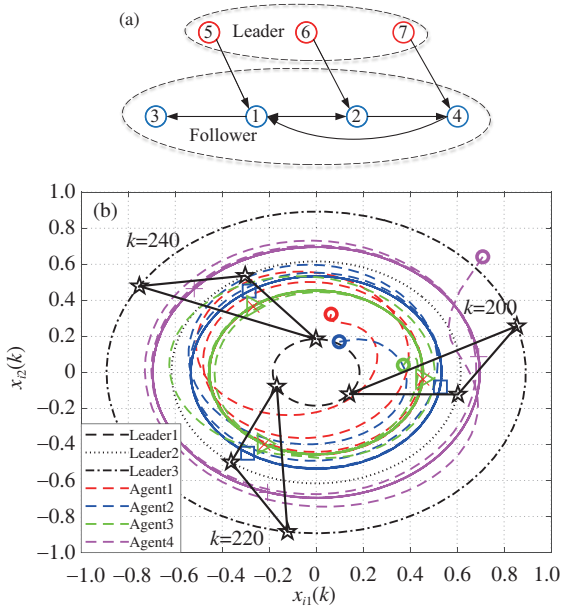

Figure 1 (Color online) (a) The leader-follower network; (b) 2-D phase plane plot.

Numerical experiments. Consider a multi-agent system with the interaction network, as shown in Figure 1(a). The nodes 1, 2, 3, 4 represent the followers and the nodes 5, 6, 7 are leaders. Figure 1(b) shows all the followers move into the interior of the convex hull formed by the leaders. Three snapshots are respectively given by marking the leaders' trajectories at $k = 200$, 220, 240, which show that the CC problem is solved. More details can be found in Appendix C.

Conclusion. We have presented a data-driven CC for a class of DT multi-agent system via reinforcement learning. A VI algorithm has been proposed to compute the optimal solution of the coupled DT-HJB equation. Then an actor-critic network has been used to compute the iterative solutions in an online learning manner using only the input and output data. Simulation results have validated the proposed VI-based containment control.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61473061, 71503206, 61104104) and Program for New Century Excellent Talents in University (Grant No. NCET-13-0091).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Hu J P, Yuan H W. Collective coordination of multi-agent systems guided by multiple leaders. Chin Phys B, 2009, 18: 3777–3782
2 Cao Y C, Ren W, Egerstedt M. Distributed containment control with multiple stationary or dynamic leaders in fixed and switching directed networks. Automatica, 2012, 48: 1586–1597
3 Chen F, Ren W, Lin Z L. Multi-leader multi-follower coordination with cohesion, dispersion, and containment control via proximity graphs. Sci China Inf Sci, 2017, 60: 110204
4 Wen G H, Zhao Y, Duan Z S, et al. Containment of higher-order multi-leader multi-agent systems: a dynamic output approach. IEEE Trans Autom Control, 2016, 61: 1135–1140
5 He W L, Qian F, Lam J, et al. Quasi-synchronization of heterogeneous dynamic networks via distributed impulsive control: error estimation, optimization and design. Automatica, 2015, 62: 249–262