论文

基于进化势博弈的多无人机传感器网络 K -覆盖孙昌浩^{①②}, 段海滨^{①*}

① 北京航空航天大学飞行器控制一体化技术重点实验室, 北京 100183;

② 中国空间技术研究院钱学森空间技术实验室, 北京 100094

* E-mail: hbduan@buaa.edu.cn

收稿日期: 2015-12-29; 接受日期: 2016-05-04; 网络出版日期: 2016-08-16

国家自然科学基金重点项目(批准号: 61333004)、国家杰出青年科学基金项目(批准号: 61425008)和国家自然科学基金面上项目(批准号: 61273054)资助

摘要 针对多无人机传感器网络 K -覆盖问题, 提出了基于势博弈与 Log-linear 学习的分布式最优传感器配置方法. 介绍了传感器最优配置的国内外研究现状, 特别对基于进化势博弈理论的方法进行了详细的阐述. 利用博弈理论对无人机分布式传感器配置问题进行建模, 通过个体局部收益函数与全局性能函数的设计将问题构造成网络势博弈. 提出了一种基于时变 Log-linear 学习的分布式求解方法, 并利用非齐次马尔可夫链理论对收敛性与最优性进行了证明. 仿真对比实验证明了所提算法的可行性、有效性和优越性.

关键词 无人机, 传感器网络, K -覆盖, 势博弈, Log-linear, 马尔可夫链理论

1 引言

无人机(unmanned aerial vehicle, UAV)是一种可利用无线电遥控设备或机载程序控制、能携带多种任务设备、执行多种任务的不载人飞行器. 因其设计上的经济性、作战使用的灵活性与军民用途的广泛性, 无人机及其技术对未来国家安全与经济发展具有重大意义. 无人机传感器网络是由大量装载结构简单、价格低廉传感器设备的无人机节点组成的网络, 在军事和民用领域发挥着越来越重要的作用, 如军情探测和环境监测等^[1]. 为获得尽可能详细的探测信息, 通常做法是配置尽可能多的传感器以覆盖整个目标区域. 然而, 在某些情况下由于环境的

限制, 上述配置方案无法实现, 因此只能将传感器安置在特定的位置或在目标区域随机分布. 此外, 能源供给方面的限制也决定了各传感器节点仅具有有限的寿命和使用时长. 因此, 多传感器配置的关键问题在于如何在保证探测效率(探测精度、覆盖面积)的前提下提高整个传感器网络的使用寿命, 这一问题受到了越来越多的关注.

为了在上述两个具有冲突的目标之间实现折中, 工程领域提出对目标区域进行冗余覆盖的解决方案, 即 K -覆盖问题. 该方案将传感器网络节点分为 K 组, 每一时间内仅有一组传感器被激活, 而其余节点则保持关闭从而节省能量, 最终实现在保证整个网络最小使用时长的基础上提高总体有效覆盖面积的目的.

引用格式: 孙昌浩, 段海滨. 基于进化势博弈的多无人机传感器网络 K -覆盖. 中国科学: 技术科学, 2016, 46: 1016–1023

Sun C H, Duan H B. An evolutionary potential game theoretic approach for the K -COVER problem in multi-UAV sensor networks (in Chinese). Sci Sin Tech, 2016, 46: 1016–1023, doi: 10.1360/N092015-00317

的. 研究表明, 某些型号的传感器在分段使用时其寿命可达到连续使用条件下的 2 倍^[2]. 本质上, 传感器最优配置问题是 K -覆盖集问题在工程领域中的实际应用.

国内外学者对传感器配置以及规划问题进行了前期的研究工作, 其中在覆盖性能的限制下最大化传感器网络使用时长(即覆盖集个数 K)的一类问题称为最大集覆盖问题. Slijepceiv 和 Potkonjak^[3]对多传感器网络最大集覆盖问题进行了研究, 证明了其是一类非确定性多项式完全(non-deterministic complete, NPC)问题, 并提出了一种计算全覆盖集的启发式算子. Bai 等人^[4]研究了异质传感器网络中可达到全覆盖与 K 联通的传感器配置模式, 并提出了一种新型的多边形配置方法. 然而, 对实际工程应用而言全覆盖的要求过于苛刻, 即要求每个传感器子集可完全覆盖感兴趣的目标区域. 因此, 为了更好地解决实际工程场景中的问题, Zhang 和 Hou^[5]提出了 α -时长的概念, 即比例为 α 的目标区域被覆盖的时间长度. 随后, Zhang 和 Hou^[6]提出了一种最大化能源受限传感器网络 α -时长的集中式算法. Jia 等人^[7]利用改进的多目标遗传算法对无线传感器网络中的覆盖问题进行了研究, 该算法在保证传感器网络全覆盖性的同时降低了系统的能源损耗.

斯坦福大学的 Abrams 等人^[8]对传感器网络 K 覆盖集问题分布式求解算法进行了首次研究, 并提出了 3 种不同的近似算法: 随机算法、分布式贪婪算法以及集中式贪婪算法. 在随机算法中, 各传感器个体随机选择一个覆盖集. 该算法简单易行, 具有很强的鲁棒性, 但覆盖效率很低. 在分布式贪婪算法中, 各传感器节点按照编号顺序依次作决策, 即选择所要加入的覆盖集. 根据之前节点策略, 各节点选择最大化自身收益的覆盖集. 虽然该算法同样简单易行, 但仍然存在着以下问题: (1) 由于各节点仅作一次决定, 该算法无法保证解的最优性; (2) 各节点在作决策时需要之前所有节点的决策信息, 因此该算法不属于严格意义上的分布式算法.

本文针对多无人机传感器网络 K -覆盖问题, 提出了基于势博弈与 Log-linear 学习的分布式最优传感器配置方法. 利用博弈理论对分布式个体间的交互过程进行建模, 通过个体局部收益函数与全局性能函数的设计将问题构造为网络势博弈. 值得注意的是, 虽然现存学习算法可对势博弈问题进行求解, 然

而大部分算法仅可保证得到一组随机的纳什均衡解. 由于势博弈可能存在不止一个纳什均衡, 其中仅有一部分纳什均衡对应于问题的全局最优解, 因此大部分算法无法保证解的全局最优性. 本文针对此问题, 基于时变 Log-linear 学习理论, 提出了面向分布式与动态环境的最优协调算法, 并利用非齐次马尔可夫链理论对其收敛性与最优性进行了证明. 仿真对比实验证明所提算法的可行性、有效性和优越性.

2 数学模型

2.1 无人机传感器网络 K 覆盖模型

n 架无人机对二维平面上某一目标区域 $\mathfrak{R}$ 进行持续定点监视与侦测, 每架无人机三维位置坐标为 $X_i=(x_i, y_i, z_i)$. 每架无人机均搭载相同的光学传感器, 并通过无线网络与周围邻居进行双向通信以获得位置与决策信息. 为保证无人机的安全飞行, 将无人机随机配置在目标区域上的三维空域, 高度范围为 $[H_{\min}, H_{\max}]$. 令分布于高度为 $H_{\max}$ 空域的无人机传感器视场半径为 $R_{\max}$, 则高度为 $z_i \in [H_{\min}, H_{\max}]$ 的无人机传感器探测半径 R_i 为

$$R_i = \frac{R_{\max} z_i}{H_{\max}}. \quad (1)$$

无人机 i 所搭载传感器视场(field of view, FOV)半径为 R_i , 通信半径为 C_i . 为延长无人机传感器网络的使用寿命, 同时保证侦查效率, 即任一时刻目标区域的有效覆盖率, 将探测时间平均分为 k 段, 并将 n 架无人机分成 k 个互不相交的覆盖集 $K=\{1, 2, \dots, k\}$. 在每一探测时段 $j \in K$, 第 j 个覆盖集中的传感器被激活, 而其余无人机上的传感器保持静默状态以节省能量. 令 $N=\{1, 2, \dots, n\}$, 无人机 i 所属集合为 $S_i \in K$, 则 $S=\{S_i | i \in N, S_i \in K\}$ 称为一个覆盖集划分.

为最大限度地提高多无人机传感器网络的有效探测面积, 覆盖集划分的目标为最大化如下目标函数:

$$f(S) = \sum_{j=1}^k f_j(S), f_j(S) = \cup \mathfrak{R}_i, S_i = j, \quad (2)$$

式中, $f(S)$ 为全局有效覆盖面积, $f_j(S)$ 表示第 j 个覆盖集的有效侦查面积, 即同属覆盖集 j 的无人机侦查区域 $\mathfrak{R}_i$ 的并集. 对任一无人机个体 i , 定义个体收益为与覆盖集 S_i 中其他无人机不相交的探测区域面积, 即

$$p_i(S) = \mathfrak{R}_i \setminus \cup(\mathfrak{R}_i \cap \mathfrak{R}_j), j \neq i \text{ 且 } S_j = S_i, \quad (3)$$

式中, “ $\setminus$ ”表示集合的减操作, $\cup(\mathfrak{R}_i \cap \mathfrak{R}_j)$ 是无人机 i 与覆盖集 j 中其余无人机重合的探测区域. 如图 1 所示, 图 1(a)中的灰色区域表示整个覆盖集 j 的探测区域面积, 即总体收益 $f(S)$, 而图 1(b)中除空白区域外的有色区域分别为各无人机个体的个体收益 $p_i(S)$.

2.2 博弈模型

博弈论是研究相互影响、互相依赖的理性个体决策行为以及均衡结果的理论. 有限策略博弈包含以下 3 个要素:

(1) 局中人集合 $N = \{1, 2, \dots, n\}$, 定义了参与博弈的所有个体;

(2) 个体行动集 $A = \{A_i | i \in N\}$, 定义了局中人可选择的全部行为, 其中的元素被称为行动, 是参与者在博弈过程中的决策变量;

(3) 支付函数 $U_i: A \rightarrow \mathbb{R}$. 其中, $A := \times_{j \in N} A_j$ 称为策略组合集, $a = (a_1, a_2, \dots, a_n) \in A$ 称为一个策略组合^[9];

定义 1^[10] 若博弈 $G = \{N, A, \{U_i\}_{i=1}^n\}$ 中任一主体的收益值 U_i 均只依赖于一部分博弈参与者 $\Omega_i \subset N$, 则此博弈称为空间网络博弈, 简称网络博弈或图博弈.

定义 2 $G = \{N, A, \{U_i\}_{i=1}^n\}$ 是具有权值的势博弈, 当且仅当存在函数 $\phi: A \rightarrow \mathbb{R}$ 满足: 对任意博弈参与者 $i \in N$, 任两个不同行动 a_i^1, a_i^2 , 以及任意 $a_{-i} = (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n) \in \times_{j \neq i} A_j$,

$$U_i(a_i^1, a_{-i}) - U_i(a_i^2, a_{-i}) = \omega_i [\phi(a_i^1, a_{-i}) - \phi(a_i^2, a_{-i})], \quad (4)$$

式中, ϕ 为势函数. 当且仅当 $\omega_i = 1, i \in N$ 时, 该博弈

图 1 (网络版彩图)总体收益与个体收益

(a) 全局收益; (b) 个体收益

称为精确势博弈, 简称势博弈.

定理 1 若各无人机的通信距离 C_i 满足 $C_i \geq \sqrt{(H_{\max} - H_{\min})^2 + (2R_{\max})^2}$, 其中 $H_{\max}$ 与 $H_{\min}$ 分别表示无人机网络中无人机个体的最大和最小高度, $R_{\max}$ 为处于高度为 $H_{\max}$ 空域无人机的探测半径, $N = \{1, 2, \dots, n\}$, $A := \times_{j \in N} A_j$, $A_j = \{1, 2, \dots, k\}$ 且各无人机支付函数 $U_i(a) = p_i$ 如式(3)所示, 则:

(1) 博弈 $G = \{N, A, U\}$ 是网络博弈;

(2) 博弈 $G = \{N, A, U\}$ 是势博弈, 势函数 $\phi(a) = f(a)$.

证明 (1) 因 $p_i(a) = \mathfrak{R}_i \setminus \cup(\mathfrak{R}_i \cap \mathfrak{R}_j), j \neq i, a_j = a_i$, 所以个体 i 的收益仅取决于与其探测区域有重叠部分的邻域无人机个体, 即

$$N_i = \{j | \|X_i^{1,2} - X_j^{1,2}\| \leq R_i + R_j\}, \quad (5)$$

$$X_i^{1,2} = (x_i, y_i). \quad (6)$$

此外, 由于各无人机通信距离

$$C_i \geq \sqrt{(H_{\max} - H_{\min})^2 + (2R_{\max})^2}. \quad (7)$$

因此, 无人机 i 的所有邻居均处于其通信范围之内, 也即, 若 $\|X_i^{1,2} - X_j^{1,2}\| \geq C_i$, 则 $\mathfrak{R}_i \cap \mathfrak{R}_j = \emptyset$. 综上所述可得, 博弈模型 G 属于网络博弈.

(2) $\forall i \in N, \forall a_i^1, a_i^2 \in A, a_i^1 \neq a_i^2$, 且

$$a_{-i} = (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n) \in \times_{j \neq i} A_j,$$

易得

$$\begin{aligned} & \phi(a_i^1, a_{-i}) - \phi(a_i^2, a_{-i}) \\ &= \sum_{j=1}^k f_j(a_i^1, a_{-i}) - \sum_{j=1}^k f_j(a_i^2, a_{-i}) \\ &= U_i(a_i^1, a_{-i}) - U_i(a_i^2, a_{-i}). \end{aligned} \quad (8)$$

因此, $G = \{N, A, U\}$ 是势博弈, 势函数 $\phi(a) = f(a)$.

定理 2 多无人机传感器协同覆盖问题(2)最优解是 K -覆盖集博弈 $G = \{N, A, U\}$ 的一个纳什均衡.

证明 反证法. 令 $a^* = (a_1^*, a_2^*, \dots, a_n^*)$ 是最优覆盖问题(2)的解. 假设 a^* 不是 $G = \{N, A, U\}$ 的纳什均衡, 则存在 a_i 满足: $U_i(a_i^*, a_{-i}^*) < U_i(a_i, a_{-i}^*)$. 又因 $G = \{N, A, U\}$ 是势博弈, 所以

$$f(a_i^*, a_{-i}^*) = \phi(a_i^*, a_{-i}^*) < \phi(a_i, a_{-i}^*) = f(a_i, a_{-i}^*).$$

因此, a^* 不是问题(2)的最优解. 此结论与假设

矛盾, 故 a^* 是 $G=\{N, A, U\}$ 的纳什均衡.

定理 3 多无人机传感器 K 覆盖集博弈 $G=\{N, A, U\}$ 至少存在一个纯纳什均衡策略.

证明 因排列组合问题(2)存在最优解, 且其最优解对应于博弈 $G=\{N, A, U\}$ 的一个纳什均衡, 所以博弈 $G=\{N, A, U\}$ 至少存在一个纳什均衡.

3 基于 Log-linear 学习的分布式算法

由于第 2 节中网络势博弈 $G=\{N, A, \{U_i\}_{i=1}^n\}$ 存在纳什均衡且至少有一个纳什均衡是分布式优化问题(2)的最优解, 因此可基于进化网络势博弈理论对此分布式协同优化问题进行求解. 然而, 在众多纳什均衡解中仅有部分纳什均衡是该博弈势函数的最优解, 即最优纳什均衡. 因此, 如何通过学习规则的设计使所有参与者策略收敛至最优纳什解是当前研究的热点与难点问题. 本节提出了一种自适应 Log-linear 学习方法, 并利用非齐次马尔可夫链理论证明了其可按概率 1 收敛至最优纳什均衡.

3.1 Log-linear 学习算法

针对网络势进化博弈 $G=\{N, A, U\}$, Log-linear 学习规则描述如下: 在每一决策时刻 t , 以相同的概率 $p_i=1/N$ 随机选择个体 $i \in N$ 进行策略更新. 在此更新过程中, 个体 i 根据玻尔兹曼分布从可行策略集 A_i 中选择策略. 具体而言, 行动 $a_i \in A_i$ 被选择的概率为

$$p_{a_i(t)} = \frac{\exp(\beta U_i(a_i, a_{-i}(t-1)))}{\sum_{\tilde{a}_i \in A_i} \exp(\beta U_i(\tilde{a}_i, a_{-i}(t-1)))}, \quad (9)$$

式中, $\beta > 0$ 为玻尔兹曼分布参数, 它反映了个体对次优行动的偏好程度, $a_{-i}(t-1)$ 表示 $t-1$ 时刻除个体 i 外所有个体的策略组合. 当 $\beta=0$ 时, 所有行动具有相同的被选概率 $p=1/|A_i|$, 而当 $\beta \rightarrow +\infty$ 时, 个体 i 从最佳应对集 $\{a_i \in A_i \mid a_i = \max_{\tilde{a}_i \in A_i} U_i(\tilde{a}_i, a_{-i}(t-1))\}$ 中随机选择行动, 此时该算法转化为最佳应对学习算法.

上述策略更新规则定义了一个马尔可夫过程, 其中所有个体策略组合 $A := \times_{j \in N} A_j$ 为状态空间, t 时刻所有个体的策略组合 $a(t) = (a_i(t))_{i \in N}$ 为一个马尔可夫状态.

定理 4 对网络势进化博弈 $G=\{N, A, U\}$ 与常值参

数 $\beta > 0$ 而言, Log-linear 学习过程是一个齐次马尔可夫链 $M_\beta = \{\{a(t)\}, A, P\}$. 其中, $a(t) = \{a_i(t)\}_{i \in N}$ 是状态变量, $A := \times_{j \in N} A_j$ 是状态空间, 状态转移矩阵 P 定义如下:

$$P(a_1, a_2) = \frac{1}{N} \sum_{i=1}^N p_{a_2^i} \cdot \mathbf{1}(a_1^i = a_2^i), \quad (10)$$

式中, 若 $a_1^i = a_2^i$, 则 $\mathbf{1}(a_1^i = a_2^i) = 1$, 否则 $\mathbf{1}(a_1^i = a_2^i) = 0$.

证明 由式(10)可得, 转移矩阵 P 仅取决于玻尔兹曼系数 β 而与时间 t 无关. 因此, 当 β 为常值时, M_β 是一个时间齐次的马尔可夫链.

定理 5^[11] 在网络进化势博弈 $G=\{N, A, U\}$ 中, 若所有个体均遵循式(9)中的学习规则, 则马尔可夫链 $M_\beta = \{\{a(t)\}, A, P\}$ 具有唯一的稳态分布

$$\mu(a) = \frac{\exp(\beta \phi(a))}{\sum_{\tilde{a} \in A} \exp(\beta \phi(\tilde{a}))}, \quad (11)$$

式中, $\mu(a)$ 是 $t \rightarrow \infty$ 时 $a(t) = a$ 的概率. 该特性称为 Log-linear 学习算法的随机稳定性.

由静态分布式(10)可得, 当玻尔兹曼常数 $\beta \rightarrow \infty$ 时,

$$\mu(a) = \begin{cases} 1/|A_{\max}|, & \text{if } a \in A_{\max}, \\ 0, & \text{else,} \end{cases} \quad (12)$$

式中, $A_{\max} = \{\tilde{a} \in A \mid \tilde{a} = \max_{\tilde{a} \in A} \phi(\tilde{a})\}$ 为最优纳什均衡解集, $|A_{\max}| \geq 1$ 为最优解集的势. 因此, 当玻尔兹曼常数 β 足够大时, 马尔可夫过程 M_β 将以任意大的概率收敛于最优解.

3.2 基于变玻尔兹曼参数的 Log-linear 学习算法

由定理 5 可得, 当 $\beta \rightarrow \infty$ 时, Log-linear 学习算法以任意大率收敛至最优解. 然而由静态分布式(11)可知, 对于任意有限的 β , 该算法总存在正的概率 $p > 0$ 收敛至非最优纳什均衡解. 因此, 若要解决此问题, 必须使玻尔兹曼分布参数 β 与更新时间 t 同步趋于无穷, 即 β 应为时间 t 的非递减函数.

3.2.1 非齐次马尔可夫链

考虑马尔可夫过程 $M = \{\{a(t)\}, A, P_t\}$, 其中 $a(t) = \{a_i(t)\}_{i \in N}$, $a_i(t) \in A_i$ 是状态变量, $A := \times_{j \in N} A_j$ 是状态空间, P_t 是状态转移矩阵. 若 P_t 随时间 t 变化, 则 M 为非齐次马尔可夫链. 令 $P_{t,n} = \prod_t^{t+n-1} P_t$, 并记 t 时

刻 n 步转移矩阵第 i 行第 j 列元素为 $p_{i,j}^{t,n}$.

定义 3^[12] 若 $\forall t, i, i', j, \lim_{n \rightarrow \infty} |p_{i,j}^{t,n} - p_{i',j}^{t,n}| = 0$, 则马尔可夫链 $M = \{\{a(t)\}, A, P_t\}$ 是弱可遍历的. 若 M 弱可遍历且 $\forall t, \lim_{n \rightarrow \infty} P_{t,n} = P_\infty$, 则称其为强可遍历.

定理 6 若非齐次马尔可夫链 $M = \{\{a(t)\}, A, P_t\}$ 满足如下条件:

- (1) $M = \{\{a(t)\}, A, P_t\}$ 是弱可遍历的;
- (2) 在每一时刻 t , 状态转移矩阵 P_t 均为对应一个齐次马尔可夫过程且具有唯一稳态分布 μ_t ;
- (3) 稳态分布 μ_t 满足

$$\sum_{t=1}^{\infty} \sum_{a \in A} |\mu_t^a - \mu_{t+1}^a| < \infty, \quad (13)$$

则称非齐次马尔可夫链 M 是强可遍历的.

3.2.2 时变 Log-linear 算法

定理 6 与强可遍历非齐次马尔可夫链的稳态分布特性表明, 可通过选择合适的玻尔兹曼参数 $\beta(t)$ 构造强可遍历非齐次马尔可夫过程 $M_{\beta(t)} = \{\{a(t)\}, A, P_t\}$, 使其稳态分布权值均集中于可最大化势函数的解上, 从而实现当时间 t 趋于无穷时算法以概率 1 收敛于最优纳什解集的目的. 此时, 在每一决策时刻行动 $a_i \in A_i$ 被选择的概率为

$$p_{a_i(t)} = \frac{\xi^{\beta(t)U_i(a_i, a_{-i}(t-1))}}{\sum_{\tilde{a}_i \in A_i} \xi^{\beta(t)U_i(\tilde{a}_i, a_{-i}(t-1))}}. \quad (14)$$

当 $\xi = e$ 时, 该分布为标准玻尔兹曼分布.

定理 7 在网络势博弈 $G = \{N, A, U\}$ 中, 若 $0 \leq U_i \leq 1, i \in N$, 且

$$\beta(t) = \frac{\log_{\xi}(\lambda t + 1)}{c}, \quad (15)$$

$$\lambda \geq 1, \quad (16)$$

$$\forall t > 0,$$

$$\begin{aligned} sp(P_{t,c}) &= sp\left(\prod_t^{t+c-1} P_t\right) \\ &= \min_{i,j} \sum_j \min(p_{i,j}^{t,c}, p_{i,j}^{t,c}) > 0, \end{aligned} \quad (17)$$

则 Log-linear 算法是一条弱可遍历马尔可夫链.

证明 将式(15)代入式(14), 得

$$p_{a_i(t)} = \frac{(\lambda t + 1)^{\frac{U_i(a_i, a_{-i}(t-1))}{c}}}{\sum_{\tilde{a}_i \in A_i} (\lambda t + 1)^{\frac{U_i(\tilde{a}_i, a_{-i}(t-1))}{c}}}. \quad (18)$$

因 $\lambda t + 1 > 1, c > 0, (\lambda t + 1)^x$ 是 x 的增函数, 且 $U_i(a_i, a_{-i}) \leq 1$, 故

$$(\lambda t + 1)^{\frac{U_i(a_i, a_{-i}) - U_i(\tilde{a}_i, a_{-i})}{c}} < (\lambda t + 1)^{1/c}, \quad (19)$$

所以,

$$\begin{aligned} p_{a_i(t)} &= \frac{1}{\sum_{\tilde{a}_i \in A_i} (\lambda t + 1)^{\frac{U_i(\tilde{a}_i, a_{-i}(t-1)) - U_i(a_i, a_{-i}(t-1))}{c}}} \\ &\geq \frac{1}{\sum_{\tilde{a}_i \in A_i} (\lambda t + 1)^{\frac{1}{c}}} \geq \frac{1}{A(\lambda t + 1)^{\frac{1}{c}}}, \end{aligned} \quad (20)$$

$$p_{a_i(t+c)} \geq \frac{1}{\alpha(\lambda t + \lambda c + 1)^{\frac{1}{c}}}, \quad (21)$$

式中, $A = \max_{i \in N} |A_i|$.

由式(10)得:

$$\begin{aligned} p_{a_1, a_2}^{t,1} &= P_t(a_1, a_2), \\ p_{a_1, a_2}^{t,2} &= \sum_{b_1 \in A} P_t(a_1, b_1) P_{t+1}(b_1, a_2), \\ p_{a_1, a_2}^{t,3} &= \sum_{b_1 \in A} \sum_{b_2 \in A} P_t(a_1, b_1) P_{t+1}(b_1, b_2) P_{t+2}(b_2, a_2), \\ &\dots \\ p_{a_1, a_2}^{t,c} &= \sum_{b_1 \in A} \sum_{b_2 \in A} \dots \sum_{b_{c-1} \in A} P_t(a_1, b_1) \\ &\quad P_{t+1}(b_1, b_2) \dots P_{t+c-1}(b_{c-1}, a_2), \\ \min_{a_1, a_2} p_{a_1, a_2}^{t,c} &\geq [P_{t+c-1}(a_1, a_2)]^c \\ &\geq \frac{1}{N^c A^c [\lambda(t+c-1)+1]}, \end{aligned} \quad (22)$$

式中, $p_{i,j}^{t,c}$ 表示 t 时刻 c 步转移矩阵 $P_{t,c}$ 中的正值元素.

由 $sp(P_{t,c}) = sp(\prod_t^{t+c-1} P_t) = \min_{i,j} \sum_j \min(p_{i,j}^{t,c}, p_{i,j}^{t,c})$ 可知, $sp(P_{t,c})$ 大于 t 步转移矩阵 $P_{t,c}$ 每列最小值之和, 即

$$sp(P_{t,c}) \geq \sum_{j \in A} \min_i p_{i,j}^{t,c}. \quad (23)$$

又由 $sp(P_{t,c}) > 0$ 得,

$$\forall i, i \in A, \exists j \in A: p_{i,j}^{t,c} > 0 \text{ 且 } p_{i,j}^{t,c} > 0. \quad (24)$$

由式(23)到(24)可得,

$$sp(P_{t,c}) \geq \min_{i,j} p_{i,j}^{t,c}. \quad (25)$$

综合式(22)与(25), 有

$$sp(P_{t,c}) \geq \frac{1}{N^c A^c [\lambda(t+c-1)+1]}. \quad (26)$$

若将更新时间 t 按 c 为单位长度进行划分, 则

$$t_j = (j-1)c + 1, j \in \mathbb{N}^+, \quad (27)$$

$$\sum_{j=1}^{\infty} sp(P_{t_j,c}) \geq \sum_{j=1}^{\infty} \frac{1}{N^c A^c (\lambda c j + 1)}. \quad (28)$$

因无穷级数 $\frac{1}{\lambda c} \sum_{j=1}^{\infty} \frac{1}{j+1}$ 发散, $c \geq 1$, 且

$\frac{1}{\lambda c j + \lambda c} \leq \frac{1}{\lambda c j + 1}$, 根据级数的比较审敛法则可得,

正项无穷级数 $\frac{1}{N^c A^c} \sum_{j=1}^{\infty} \frac{1}{(\lambda c j + 1)}$ 发散. 综上可得,

$\sum_{j=1}^{\infty} sp(P_{t_j,c}) > \infty$. 由引理 3 得, 所考虑的马尔可夫链

是弱可遍历的, 其中 $n_i = c, i \in \mathbb{N}^+$.

基于自适应 Log-linear 学习规则下马尔可夫链的弱可遍历特性与强可遍历马尔可夫链的充要条件, 有如下结论.

定理 8 设 $G=\{N,A,U\}$ 是有 N 个参与者的网络势博弈, 其中 ϕ 为势函数, $N = \{1, 2, \dots, n\}$ 为参与者集合, $A = \{A_i | i \in N\}$ 为个体策略组合集合, $a = (a_1, a_2, \dots, a_n) \in A$ 称为一个策略组合, U_i 为收益函数且 $0 \leq U_i \leq 1, i \in N$. 若博弈参与者均遵循 $\beta(t) = \frac{\log_{\xi}(\lambda t + 1)}{c}$ 的自适应 Log-linear 学习规则, 则该算法以概率 1 收敛于最大化势函数的解.

4 仿真实验及结果对比分析

为验证所提非齐次 Log-linear 学习算法在解决多无人机传感器分布式协同控制问题中的有效性与优势, 在不同场景下进行仿真验证, 并与现存算法进行了结果对比与分析, 包括最佳应对学习算法与参数固定的齐次 Log-linear 学习算法.

因每架无人机均搭载相同的光学传感器, 故当

所有无人机具有相同高度时, 该问题是一个同质传感器协同分配问题, 即所有无人机传感器在目标区域具有相同的探测视场. 表 1 与 2 分别给出了算法仿真参数设置与各无人机个体的初始三维位置分布, 其探测区域位置分布如图 2 所示.

表 1 仿真参数设置

变量符号	物理含义	变量值
n	无人机数量	10
$H_{\max}$	无人最高高度	0.5 km
$R_{\max}$	无人机最大探测半径	1 km
C	通讯距离	2 km
k	协同探测分组数	2
Monte	蒙特卡罗仿真次数	50

表 2 无人机三维位置分布

序号	1	2	3	4	5
$x(\text{km})$	2.0553	3.9495	1.3908	1.8225	3.1894
$y(\text{km})$	1.9275	2.7669	3.4047	2.8260	3.5957
$z(\text{km})$	0.5	0.5	0.5	0.5	0.5
序号	6	7	8	9	10
$x(\text{km})$	2.3755	3.5411	3.9617	3.6211	2.2124
$y(\text{km})$	3.8417	1.9714	3.9007	2.5600	3.1989
$z(\text{km})$	0.5	0.5	0.5	0.5	0.5

图 3 给出了最佳应对学习算法与所提算法在 50 次蒙特卡罗仿真中的平均结果. 从图中可以看出, 最佳应对算法以很快的速度收敛于某一纳什均衡解, 平均收敛时间 $T_c < 50$ 代; 平衡状态具有强稳定性, 即每当算法收敛至任一纳什均衡状态, 则该状态将永远保持, 永不变动. 上述结果均归因于该算法所具有的“贪婪性”与“短视性”, 即每一个体在做决策时仅以概率 1 选择当前意义上最优的策略, 而其他策略被选概率均为 0.

图 3(b)为所提算法在 50 次蒙特卡罗仿真实验中的平均结果, 实验中参数设置如下: 个体收益上确界 $M=\pi(R_{\max})^2=\pi$, $\varepsilon=10 \times 10^{-10}$, $c=6.9315 \times 10^{10}$, $\lambda=c^3$, $\zeta=10 \times 10^{-11}$, 不同于最佳应对学习算法中的确定性过程, 非齐次 Log-linear 算法对应于一个随机过程. 因此, 虽然非齐次 Log-linear 算法已被证明当迭代次数 t 趋于无穷时必收敛于最优解, 但其单次仿真实验进化曲线并不具有与最佳应对学习算法类似的严格非递减与超快速收敛特性. 然而, 从统计意义上讲, 该算法却有着现存确定性算法无法比拟的优点, 即进化结果的最优性. 从对比结果可得如下结论:

图2 (网络版彩图)传感器探测区域分布

(a) 三维视图; (b) 平视图

图3 (网络版彩图)平均覆盖面积进化曲线

(a) 最佳应对算法; (b) 自适应 Log-linear

(1) 非齐次 Log-linear 算法可在较短的学习时间内得到令人满意且优于现存算法的传感器分组结果, 其中 $T_{\max}=1000$ 时所得解与最优解误差仅为 3.9%;

(2) 平均进化曲线具有近似非递减特性, 即全局有效覆盖面积随着时间 t 的增加而呈现出阶段性上升的趋势;

(3) 非齐次 Log-linear 学习算法不会在短时间内稳定于某一状态, 只有在 t 趋于无穷时严格收敛于最优解, 因此在实际应用中存在着精度与算法运行时间的权衡与折中。

5 结论

本文以多无人机协同监视与探测任务中 K -覆盖

问题为背景, 利用网络进化势博弈理论对多无人机分布式传感器协同分配问题进行了研究. 首先, 对网络势博弈理论中的主要概念进行了介绍, 并基于此理论对多无人机分布式传感器协同分配问题建立了博弈模型. 通过无人机个体局部收益函数与全局性能函数的设计, 将分布式优化问题构造为网络势博弈模型. 其次, 提出了非齐次 Log-linear 分布式学习算法, 并利用非齐次马尔科夫链理论对算法的收敛性与最优性进行了分析与证明. 理论分析表明, 该算法可在收敛时间 t 趋于无穷时以概率 1 收敛于最优纳什均衡解. 最后, 利用仿真实验对所提方法进行了验证, 并与最佳应对策略学习算法进行了对比分析. 仿真结果表明, 所提算法可在较短的学习时间内提高整个无人机网络的有效覆盖面积。

参考文献

- 1 Duan H B, Li P. Bio-inspired Computation in Unmanned Aerial Vehicles. Berlin, Heidelberg: Springer-Verlag, 2014

- 2 Benini L, Castelli G, Macii A, et al. A discrete-time battery model for high-level power estimation. In: Proceedings of the Conference on Design, Automation and Test in Europe. ACM New York, 2000. 35–41
- 3 Slijepcevic S, Potkonjak M. Power efficient organization of wireless sensor networks. In: IEEE International Conference on Communications. Helsinki, 11 Jun 2001–14 Jun 2001. 2001, 2: 472–476
- 4 Bai X, Xuan D, Yun Z, et al. Complete optimal deployment patterns for full-coverage and k -connectivity ($k \leq 6$) wireless sensor networks. In: Proceedings of the 9th ACM International Symposium on Mobile ad Hoc Networking and Computing, 2008. 401–410
- 5 Zhang H, Hou J C. On the upper bound of α -lifetime for large sensor networks. ACM T Sensor Netw, 2005, 1: 272–300
- 6 Zhang H, Hou J C. Maximising α -lifetime for wireless sensor networks. Int J Sensor Netw, 2006, 1: 64–71
- 7 Jia J, Chen J, Chang G R, et al. Efficient cover set selection in wireless sensor networks. Acta Automatica Sin, 2008, 34: 1157–1162
- 8 Abrams Z, Goel A, Plotkin S. Set k -cover algorithms for energy efficient monitoring in wireless sensor networks. In: Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks. ACM, 2004. 424–432
- 9 张国鹏. 基于博弈论的无线网络资源竞争与协作机制研究. 博士学位论文. 西安: 西安电子科技大学, 2009
- 10 Kearns M, Littman M L, Singh S. Graphical models for game theory. In: Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 2001. 253–260
- 11 Blume L E. The statistical mechanics of strategic interaction. Games Econ Behav, 1993, 5: 387–424
- 12 Isaacson D L, Madsen R W. Markov chains, theory and applications. In: Wiley Series in Probability and Mathematical Statistics. New York: Wiley, 1976

An evolutionary potential game theoretic approach for the K -COVER problem in multi-UAV sensor networks

SUN ChangHao^{1,2} & DUAN HaiBin¹

¹ Bio-inspired Autonomous Flight Systems Research Group, Science and Technology on Aircraft Control Laboratory, Beihang University Beijing 100191, China;

² Qian Xuesen Laboratory of Space Technology, China Academy of Space Technology, Beijing 100094, China

This paper focuses on the K -COVER problem in unmanned aerial vehicle (UAV) sensor networks and proposes a game theoretic approach based on the network potential game and Log-linear learning. After detailing the research progress at home and abroad, we firstly model the K -COVER problem in UAV networks as network potential game via appropriately designing of individual and global payoff functions. Secondly, a Log-linear learning based algorithm is proposed, whose convergence and optimality are proven using inhomogeneous Markov chain theory. Simulation results demonstrate the feasibility, effectiveness, and superiority.

unmanned aerial vehicle (UAV), sensor network, K -COVER, evolutionary game theory, Log-learning, Markov chain

doi: 10.1360/N092015-00317