

doi: 10.7690/bgzdh.2020.07.011

基于拥塞博弈的多无人机自主侦察任务规划

赵玉亮¹, 宋业新¹, 康丽文²

(1. 海军工程大学基础部, 武汉 430033; 2. 海图信息中心, 天津 300450)

摘要: 为解决通信中断时多无人机自主侦察任务规划的问题, 提出一种基于拥塞博弈的分布式任务规划方法。根据问题场景给出各无人机的侦察收益函数以及各子区域的社会福利函数, 基于拥塞博弈, 建立多无人机自主侦察拥塞博弈模型, 验证纯纳什均衡的存在性, 给出多无人机自主侦察拥塞博弈的无政府状态代价和稳定状态代价的界限, 设计一种具有性能保证并且易于实现的策略更新规则, 并通过蒙特卡罗方法进行模拟仿真实验。仿真结果验证了该模型的适用性和策略更新规则的有效性。

关键词: 多无人机; 自主侦察; 通信中断; 拥塞博弈; 纳什均衡; 无政府状态代价; 稳定状态代价; 策略更新规则

中图分类号: TP24 文献标志码: A

Multi-UAV Autonomous Reconnaissance Mission Planning Based on Congestion Game

Zhao Yuliang¹, Song Yexin¹, Kang Liwen²

(1. Department of Basic, Naval University of Engineering, Wuhan 430033, China;
2. Marine Map Information Centre, Tianjin 300450, China)

Abstract: In order to solve the problem of autonomous reconnaissance mission planning of multi UAV when communication is interrupted, a distributed mission planning method based on congestion game is proposed. According to the problem scenario, the reconnaissance payoff function of each UAV and the social welfare function of each sub-region are given. Based on the congestion game, a congestion game model for autonomous reconnaissance of multi-UAV is established. Next, the existence of pure Nash equilibrium is verified. Give the boundary between the cost of anarchy and the cost of stable state of multi-UAV autonomous reconnaissance congestion game. Furthermore, a strategy updating rule with performance guarantee and easy implementation is designed and simulated by Monte Carlo method. Simulation results verify the applicability of the model and the validity of policy updating rules.

Keywords: multi UAV; autonomous reconnaissance; interruption of communication; congestion game; Nash equilibrium; price of anarchy; price of stability; strategy updating rule

0 引言

近年来,无人机的可靠性、持久性和多功能性已经得到了普遍证实^[1]。相比有人机,无人机的响应时间更短,当任务环境恶劣或者当信息收集对有人机过于危险时,无人机可以快速地捕捉任务区域中静止或运动目标的实时图像;因此,无人机在战场上执行侦察任务受到越来越多的关注^[2-4]。由于通信延迟、通信设备传输功率有限、通信带宽有限^[5-6]以及复杂通信电子环境下对无人机电子静默的要求^[7],往往导致无人机无法可靠地与其他无人机或者全局控制中心进行信息交换;因此,研究有效应对多无人机之间通信不可靠的问题,保障无人机安全高效地完成侦察任务,具有十分重要的现实意义。

现阶段,国内外对于该问题的研究尚不多见:

Sujit 等^[8]研究了一种在通信距离限制下基于谈判机制的多无人机任务分配方法,并与贪心算法进行比较,从而验证其有效性;Mirzaei 等^[9]研究了通信延迟下的多无人机协同搜索问题,使用随机预测方法补偿通信延迟对系统的影响;Bardhan 等^[10]研究了传感器范围和通信带宽的限制条件下的无人机任务分配问题,基于最大熵方法使无人机能够利用可用信息获得其他无人机策略集的主观概率,并利用相关均衡来协调它们的策略;Pei Li 等^[11]基于势博弈提出了一种多无人机协同搜索与监视的博弈论模型,并采用二叉对数线性学习进行运动控制。这些学者的研究主要针对通信约束条件展开,依赖于无人机之间的部分通信,没有考虑完全无通信条件下的无人机任务规划。

收稿日期: 2020-03-19; 修回日期: 2020-04-10

基金项目: 国家自然科学基金资助(71171198)

作者简介: 赵玉亮(1991—),男,山东人,博士,从事无人机任务规划、博弈论研究。E-mail: yl.zhao0811@gmail.com。

笔者针对多无人机在通信中断时侦察任务规划的问题，提出一种基于拥塞博弈的分布式任务规划方法，使无人机可以不依赖可靠的通信媒介与其他无人机协调而进行自主规划。

1 基本概念

定义 1 一个有限的策略式博弈，一般由局中人、策略集和收益函数 3 部分组成。笔者用 $G(N,S,U)$ 表示一个有限的策略式博弈，其中：

- 1) 局中人集合： $N=\{1,2,\dots,n\}$ ；
- 2) 策略集： $S_i=\{s_1^i,s_2^i,\dots,s_{k_i}^i\}$ ， $i\in N$ 表示第 i

个局中人的策略集，又记 $S_{-i}=\times_{j\neq i}S_j$ ，表示除了第 i 个局中人外，其余 $n-1$ 个局中人的策略组合集合；

3) 局势：把每个局中人的策略集中各取一个策略所组成的策略组成为局势，记所有局势的集合为 $S=\times_i S_i$ (“ \times ”表示笛卡尔积)；

4) 收益函数集： $U=(U_1,U_2,\dots,U_n)$ ，其中 $U_i\in R$ 为第 i 个局中人的收益函数。

注 这里的“有限”是指：1) 局中人数 $N<\infty$ ；2) 策略数 $|S_i|<\infty$ ， $i\in N$ 。

定义 2 对于一个给定的有限策略式博弈 $G(N,S,U)$ ，如果对于任意的 $i\in N$ 和任意的 $s_i\in S_i$ ，都有

$$U_i(s_i^*,s_{-i}^*)\geq U_i(s_i^*,s_{-i}^*),$$

则称局势 $s^*=(s_1^*,s_2^*,\dots,s_n^*)$ 是 G 的一个纯纳什均衡。

定义 3 社会福利是指在一个博弈结果中所有局中人个体收益的总和，即

$$W(s)=\sum_{i\in N}U_i(s).$$

为了更好地理解博弈均衡相较于系统级最优社会福利函数的效率情况，笔者将使用无政府状态代价(price of anarchy, PoA)和稳定状态代价(price of stability, PoS)来衡量博弈均衡的低效率(Inefficiency)。它们的定义如下所示：

定义 4^[12] PoA 为最差均衡的社会福利与博弈系统最优社会福利之间的比率，即

$$PoA = \min_{s^e\in E(G)} (W(s^e)/W(s^{opt})).$$

其中： $s^{opt}\in \arg \max_{s\in S} W(s)$ ； $E(G)$ 为博弈 G 的均衡集合。

定义 5^[11] PoS 为最优均衡的社会福利和博弈系统最优社会福利之间的比率，即

$$PoS = \max_{s^e\in E(G)} (W(s^e)/W(s^{opt})).$$

在以上定义中，最差(最优)均衡是指在所有均衡中，给出最小(最大)社会福利的均衡(包括纯纳什均衡、混合纳什均衡等)。

表征均衡的低效率通常涉及到非平凡领域的特定分析。下面的定义和定理给出了一种提供表征博弈均衡效率损失的通用方法。

定义 6^[13] 对于任意一个博弈 G ，如果存在参数 λ 和 $\mu>-1$ 使得任意 2 个局势 $s,s^*\in S$ ，都有

$$\sum_{i\in N}U_i(s_i^*,s_{-i})\geq \lambda\sum_{i\in N}U_i(s^*)-\mu\sum_{i\in N}U_i(s)$$

则称博弈 G 是 (λ,μ) -平滑的。

定理 1^[13] 一个 (λ,μ) -平滑的博弈的 PoA 至少为 $\lambda/(1+\mu)$ ，并且此 PoA 的界限适用于任何均衡概念(包括纯纳什均衡，混合纳什均衡等)。

笔者所建的无人机侦察任务规划模型与拥塞博弈密切相关，其定义如下：

定义 7 一个拥塞博弈 G 由以下 5 个部分组成， $(N,M,(S_i)_{i\in N},(C_j)_{j\in N})$ ，其中：

- 1) $N=\{1,2,\dots,n\}$ 为局中人集合，这里假设有 n 个局中人；
- 2) M 为资源集合；
- 3) $S_i\subseteq 2^M$ 为局中人 $i\in N$ 策略集；
- 4) $C_j:2^N\rightarrow R^+$ 为资源 $j\in M$ 的一个联合价值函数，表示一组局中人从资源 j 中得到的资源价值总和；
- 5) $U_i:R^+\rightarrow R^+$ 为局中人 $i\in N$ 的收益函数。

在一个局势 $s=(s_1,s_2,\dots,s_n)\in S$ 中，将选择资源 $j\in M$ 的局中人表示为集合 $Q(j)\subseteq N$ ，则 $Q(j)$ 中的局中人共享资源 j 后得到的总收益为 $C_j(Q(j))$ ，而局中人 $i\in N$ 从资源 j 中取得的收益为 $U_i(C_j(Q(j)))$ ，从而局中人 i 在局势 s 中取得的总收益为 $U_i(\sum_{j\in Q_j} C_j(Q(j)))$ 。

2 问题描述

在无人机侦察任务规划场景中，包括一个地面规划控制中心、一组侦察型无人机以及一个侦察区域，无人机需要通过对该区域进行侦察来收集信息。每架无人机拥有充分的自主权，既能通过不断地接收地面规划控制中心的指令来进行下一步的侦察动作，又能独立地决定下一步的侦察策略。在通信可靠的情况下，地面规划控制中心会制定一个合作的规划方案来最大限度地提高无人机编队的侦察收

益，所有的无人机需要执行地面控制中心发送的任务指令。当通信突然不可靠或者进入通信禁区时，无人机无法通过通信来获取下一步的任务指令，需要独立地作出决策。无人机的自主性和通信的缺乏，使得它们之间不能相互协调出一个合作的解决方案。在通信中断时，无人机在只知道其他无人机的侦察位置以及历史侦察路径等有限信息的情况下，如何将自己的侦察收益最大化成为研究重点。

3 多无人机自主侦察拥塞博弈模型

笔者将基于拥塞博弈来构建多无人机自主侦察任务规划模型，并求证多无人机自主侦察拥塞博弈模型的纯纳什均衡的存在性，给出多无人机自主侦察拥塞博弈模型的 PoS 和 PoA 界限。

3.1 侦察任务规划模型

首先基于拥塞博弈构建多无人机自主侦察任务规划模型。

设 $N=\{1,2,\dots,n\}$ 为无人机的集合，对应于拥塞博弈的局中人集合。侦察的地理区域被离散成一个有限子区域集 M ，对应于拥塞博弈的资源集合。侦察区域可以由有向图 D 表示，其中图的顶点代表子区域，图的有向边代表这些子区域之间的连接性。此外，无人机在子区域间的移动和侦察子区域需要花费时间。为简化模型，笔者将时间离散为时间步长，无人机在相邻子区域间移动并侦察后续子区域总共花费一个时间步长。记通信中断时，无人机还能够侦察的时间步长为 l ，则无人机 $i \in N$ 的策略集 S_i 为一组，从通信中断时的位置出发的长度为 l 的侦察路径。又由于无人机通过机载传感器收集各子区域的可用信息，而不同的无人机的传感器的信息收集能力不同；因此，笔者设无人机 i 的传感器的侦察子区域 $j \in M$ 的有效概率为 $p_i(j) \in (0,1]$ 。与此同时，每个子区域 j 都有一个信息价值，笔者记通信中断时子区域 j 的剩余信息价值为 $v(j): M \rightarrow R^+$ 。每架无人机的目标是在 l 个时间步长内最大化其侦察收益。

接下来给出子区域 $j \in M$ 的收益函数 C_j 和无人机 $i \in N$ 的收益函数 U_i 的具体公式。首先，记 $U_i(s,j)$ 为无人机 $i \in N$ 在局势 $s \in S$ 下从子区域 $j \in M$ 获取的信息价值总和。可见：在现实情形下，多架无人机对于同一子区域的侦察有时是同时进行的，有时不是同时进行的。特殊地，由于侦察任务的需要，某些无人机可能不止一次地侦察同一子区域，则同一

架无人机对于同一子区域的多次侦察存在侦察的先后次序。对于任意无人机在访问某一子区域时，其获得的侦察收益受到在其之前侦察该子区域的无人机的影响，也受到同时侦察该子区域的其他无人机的影响。

具体地，设在局势 $s \in S$ 下有一组无人机 $Q(j)$ 侦察子区域 j ，其中 $Q(j)$ 中的无人机按侦察次序依次排列。此外，由于同一无人机 i 可能多次侦察子区域 j ，则集合中可能包含多个 i ，即 $Q(j)$ 是一个多重集，例如： $Q(j)=(3,2,4,3,5,1)$ 。又假设在局势 s 下，无人机 i 侦察子区域 j 共 t 次。为准确地描述每架无人机的收益函数，笔者定义 2 个序列： $A_k^i(j), A_2^i(j), \dots, A_t^i(j)$ 与 $B_1^i(j), B_2^i(j), \dots, B_t^i(j)$ 。其中： $A_k^i(j)$ 表示与无人机 i 对子区域 j 的第 $k(1 \leq k \leq t)$ 次侦察同步无人机的集合； $B_k^i(j)$ 表示在无人机 i 对子区域 j 第 k 次侦察之前就已经侦察过子区域 j 无人机的集合。显然， $A_k^i(j)$ 必包含一个 i ，而 $B_k^i(j)$ 是一个多重集，包含 $k-1$ 个 i 。因此，无人机 i 在子区域 j 中的侦察收益可表示为：

$$U_i(s,j) = \sum_{k=1}^t \left(p_i(j) \prod_{z \in A_k^i(j)} p_z(j) (1 - \prod_{z \in B_k^i(j)} (1 - p_z(j))) v(j) \right). \quad (1)$$

$Q(j)$ 中所有无人机在子区域 j 中获取的总侦察收益为：

$$C_j(Q(j)) = (1 - \prod_{z \in Q(j)} (1 - p_z(j))) v(j). \quad (2)$$

3.2 纯纳什均衡的存在性

多无人机自主侦察拥塞博弈的纯纳什均衡并不总是存在的，下面通过实例来说明。

设有 2 架无人机 A、B，侦察区域共划分为 12 个侦察子区域，其连接性如图 1 所示。无人机 A 的初始位置为子区域 1，无人机 B 的初始位置为子区域 3，无人机 A、B 对于各点的侦察有效概率均为 1。设 $l=4$ ，子区域 1-12 的侦察信息价值分别为：0, 0, 0, 0, 10, 10, 10, 10, 20, 20, 10, 25。由图可知，无人机 A 有 2 个策略：1→2→6→10→9, 1→2→6→5→9；无人机 B 也有 2 个策略：3→4→8→12→11, 3→7→11→10→9。则该博弈的收益矩阵为：

$$\begin{bmatrix} (50,45) & (30,40) \\ (40,45) & (30,50) \end{bmatrix}。$$

很明显，在此情形下，该博弈没有纯纳什均衡。

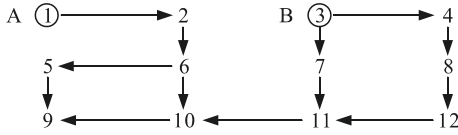


图 1 实例 1

3.3 PoA 和 PoS 界限

笔者给出多无人机自主侦察拥塞博弈的 PoA 和 PoS 界限的相关结果。

为方便起见,首先对式(1)进行简化。记 $U_{B_k^i(j)}^{A_k^i(j)}(s)$ 为无人机 i 在局势 $s \in S$ 下,在子区域 $j \in M$ 已被多重集 $B_k^i(j)$ 中的无人机侦察过的情况下,在无人机 i 的第 k 次侦察的过程中,与集合 $A_k^i(j)$ 中的无人机共同获取的侦察收益,其表达式为

$$U_{B_k^i(j)}^{A_k^i(j)}(s) = (1 - \prod_{z \in A_k^i(j)} (1 - p_z(j))) \prod_{z \in B_k^i(j)} (1 - p_z(j)) v(j)。$$

式(1)可以简记为

$$U_i(s, j) = \sum_{k=1}^t \left(\frac{p_i(j)}{\sum_{z \in A_k^i(j)} p_z(j)} U_{B_k^i(j)}^{A_k^i(j)}(s) \right)。$$

又记 $Q(j)|_i$ 为 $Q(j)$ 中无人机 i 的集合,即 $Q(j)|_i$ 中只包含 i ,并且与 $Q(j)$ 中的 i 具有相同的多重性。此外,记 $Q(j)|_{-i}$ 为 $Q(j)|_{-i} = Q(j) \setminus Q(j)|_i$,即 $Q(j)$ 除去 $Q(j)|_i$ 的剩余无人机的集合,则有如下引理成立:

引理 1 对于某个侦察子区域 $j \in M$,设多重集 $Q(j)$ 为侦察子区域 j 的无人机集合, $S' \subseteq S$ 为所有以 $Q(j)$ 为侦察子区域 j 的无人机集合的策略集,则对于 $\forall i \in Q(j)$ 及 $\forall s \in S'$,有 $U_i(s, j) \geq U_{Q(j)|_{-i}}^{Q(j)|_i}(s)$ 成立。

证明:类似于策略 $s \in S'$ 下, $A_1^i(j), A_2^i(j), \dots, A_t^i(j)$ 与 $B_1^i(j), B_2^i(j), \dots, B_t^i(j)$ 的定义,笔者定义在策略 $\bar{s} \in S'$ 下的 $\bar{A}_1^i(j), \bar{A}_2^i(j), \dots, \bar{A}_t^i(j)$ 与 $\bar{B}_1^i(j), \bar{B}_2^i(j), \dots, \bar{B}_t^i(j)$, 其与 $A_1^i(j), A_2^i(j), \dots, A_t^i(j)$ 和 $B_1^i(j), B_2^i(j), \dots, B_t^i(j)$ 的唯一区别在于,对于 $\forall k (1 \leq k \leq t)$, $\bar{A}_k^i(j)$ 只包含一个元素 i , $A_k^i(j)$ 中与 i 同步的其他无人机在策略 \bar{s} 下都先于无人机 i 的第 k 次侦察而侦察子区域 j , 即 $\bar{A}_k^i(j) = \{i\}$, $\bar{B}_k^i(j) = B_k^i(j) \cup (A_k^i(j) \setminus \{i\})$ 。显然,在局势 \bar{s} 下,侦察子区域 j 的无人机集合仍为多重集 $Q(j)$ 。根据以上定义,在策略 \bar{s} 下,无人机 i 在子区域 j 中的侦察收益为

$$U_i(\bar{s}, j) = \sum_{k=1}^t U_{\bar{B}_k^i(j)}^{\bar{A}_k^i(j)}(\bar{s})。$$

又

$$U_i(\bar{s}, j) = \sum_{k=1}^t U_{\bar{B}_k^i(j)}^{\bar{A}_k^i(j)}(\bar{s}) =$$

$$\sum_{k=1}^t \left(p_i(j) \prod_{z \in \bar{B}_k^i(j)} (1 - p_z(j)) v(j) \prod_{z \in \bar{A}_k^i(j) \setminus \{i\}} (1 - p_z(j)) \right),$$

$$\text{而 } \prod_{z \in \bar{A}_k^i(j) \setminus \{i\}} (1 - p_z(j)) \leq$$

$$\frac{1 - (1 - p_i(j)) \prod_{z \in \bar{A}_k^i(j) \setminus \{i\}} (1 - p_z(j))}{p_i(j) + \sum_{z \in \bar{A}_k^i(j) \setminus \{i\}} p_z(j)} = \frac{1 - \prod_{z \in \bar{A}_k^i(j)} (1 - p_z(j))}{\sum_{z \in \bar{A}_k^i(j)} p_z(j)},$$

$$\text{则 } U_i(\bar{s}, j) \leq \sum_{k=1}^t \left(p_i(j) \prod_{z \in \bar{B}_k^i(j)} (1 - p_z(j)) v(j) \cdot$$

$$\left(1 - \prod_{z \in \bar{A}_k^i(j)} (1 - p_z(j)) \right) / \sum_{z \in \bar{A}_k^i(j)} p_z(j) \right) = U_i(s, j)。$$

又定义策略 $\hat{s} \in S$, 同时定义 2 个序列 $\hat{A}_1^i(j), \hat{A}_2^i(j), \dots, \hat{A}_t^i(j)$ 与 $\hat{B}_1^i(j), \hat{B}_2^i(j), \dots, \hat{B}_t^i(j)$ 。在策略 \hat{s} 下,无人机 i 对子区域 j 的侦察是在其余所有无人机侦察完之后才进行的,即对于 $\forall k (1 \leq k \leq t)$, $\hat{A}_k^i(j) = \{i\}$, $\hat{B}_k^i(j) = Q(j)|_{-i} \cup ((k-1) \otimes \{i\})$, 其中 $(k-1) \otimes \{i\}$ 表示 $(k-1)$ 个 i 的集合。显然,在策略 \hat{s} 下,侦察子区域 j 的无人机集合 $\hat{Q}(j)$ 与 $Q(j)$ 具有相同的多重性, $\hat{A}_k^i(j)$ 和 $\bar{A}_k^i(j)$ 中的 i 具有相同的多重性。而 $\bar{B}_k^i(j) \subseteq \hat{B}_k^i(j)$, 则 $U_{\bar{B}_k^i(j)}^{\bar{A}_k^i(j)}(\bar{s}) \geq U_{\hat{B}_k^i(j)}^{\hat{A}_k^i(j)}(\hat{s})$, 进而:

$$U_i(\bar{s}, j) = \sum_{k=1}^t U_{\bar{B}_k^i(j)}^{\bar{A}_k^i(j)}(\bar{s}) \geq \sum_{k=1}^t U_{\hat{B}_k^i(j)}^{\hat{A}_k^i(j)}(\hat{s}) \geq$$

$$\sum_{k=1}^t U_{\hat{B}_k^i(j)}^{\hat{A}_k^i(j)}(\hat{s}) = \sum_{k=1}^t U_{Q(j)|_{-i} \cup ((k-1) \otimes \{i\})}^{\{i\}}(\hat{s}) \geq$$

$$\sum_{k=1}^t (U_{\phi}^{Q(j)|_{-i} \cup (k \otimes \{i\})}(\hat{s}) - U_{\phi}^{Q(j)|_{-i} \cup ((k-1) \otimes \{i\})}(\hat{s})) \geq$$

$$U_{\phi}^{Q(j)|_{-i} \cup (t \otimes \{i\})}(\hat{s}) - U_{\phi}^{Q(j)|_{-i}}(\hat{s}) \geq U_{t \otimes \{i\}}^{Q(j)|_{-i}}(\hat{s}) \geq$$

$$U_{Q(j)|_i}^{Q(j)|_{-i}}(\hat{s}) = U_{Q(j)|_i}^{Q(j)|_{-i}}(s)。$$

综上所述,对于 $\forall i \in Q(j)$ 及 $\forall s \in S'$,有 $U_i(s, j) \geq U_{Q(j)|_{-i}}^{Q(j)|_i}(s)$ 成立。

引理得证。

接下来,设任意 2 个策略 $s, s^* \in S$, 记 $Q(j)$ 和 $Q^*(j)$ 分别为 2 个策略下侦察子区域 j 的无人机集合,则策略 $(s_i^*, s_{-i}) = Q^*(j)|_i \cup Q(j)|_{-i}$, 又:

$$\sum_{i \in N} U_i(s) = \sum_{i \in N} \sum_{j \in M} U_i(s, j) = \sum_{j \in M} \sum_{i \in N} U_i(s, j) = \sum_{j \in M} U_\phi^{Q(j)}(s);$$

$$\sum_{i \in N} U_i(s^*) = \sum_{j \in M} U_\phi^{Q^*(j)}(s^*);$$

$$\sum_{i \in N} U_i(s_i^*, s_{-i}) = \sum_{i \in N} \sum_{j \in M} U_i((s_i^*, s_{-i}), j) =$$

$$\sum_{j \in M} \sum_{i \in N} U_i((s_i^*, s_{-i}), j) = \sum_{j \in M} \sum_{i \in Q^*(j)} U_i((s_i^*, s_{-i}), j).$$

应用引理 1，有 $\sum_{i \in N} U_i(s_i^*, s_{-i}) \geq$

$$\sum_{j \in M} \sum_{i \in Q^*(j)} U_{Q(j)l_i}^{Q^*(j)}(s_i^*, s_{-i}).$$

$$\sum_{i \in N} U_i(s_i^*, s_{-i}) + \sum_{i \in N} U_i(s) - \sum_{i \in N} U_i(s^*) \geq$$

$$\sum_{j \in M} \left(\sum_{i \in Q^*(j)} U_{Q(j)l_i}^{Q^*(j)}(s_i^*, s_{-i}) + U_\phi^{Q(j)}(s) - U_\phi^{Q^*(j)}(s^*) \right).$$

下面证明对于 $\forall j$ 都有 $\sum_{i \in Q^*(j)} U_{Q(j)l_i}^{Q^*(j)}(s_i^*, s_{-i}) +$

$$U_\phi^{Q(j)}(s) - U_\phi^{Q^*(j)}(s^*) \geq 0.$$

证明：

$$\sum_{i \in Q^*(j)} U_{Q(j)l_i}^{Q^*(j)}(s_i^*, s_{-i}) + U_\phi^{Q(j)}(s) \geq$$

$$\sum_{i \in Q^*(j) \setminus Q(j)} U_{Q(j)l_i}^{Q^*(j)}(s_i^*, s_{-i}) + U_\phi^{Q(j)}(s) =$$

$$\sum_{i \in Q^*(j) \setminus Q(j)} U_{Q(j)l_i}^{Q^*(j)}(s_i^*, s_{-i}) + U_{Q(j)l_x}^{Q^*(j)}(s_x^*, s_{-x}) + U_\phi^{Q(j)}(s) =$$

$$\sum_{i \in Q^*(j) \setminus Q(j)} U_{Q(j)l_i}^{Q^*(j)}(s_i^*, s_{-i}) + U_{Q(j)l_x}^{Q^*(j)}(s_x^*, s_{-x}) + U_\phi^{Q(j)}(s) \geq$$

$$U_{Q(j)l_x}^{Q^*(j)}(s_x^*, s_{-x}) + U_{Q(j)l_x}^{Q^*(j)}(s_x^*, s_{-x}) + U_\phi^{Q(j)}(s) =$$

$$\sum_{i \in Q^*(j) \setminus Q(j)} U_{Q(j)l_i}^{Q^*(j)}(s_i^*, s_{-i}) + U_{Q(j)l_x}^{Q^*(j)}(s_x^*, s_{-x}) + U_\phi^{Q(j)}(s) =$$

$$\sum_{i \in Q^*(j) \setminus Q(j)} U_{Q(j)l_i}^{Q^*(j)}(s_i^*, s_{-i}) + U_{Q(j)l_x}^{Q^*(j)}(s_x^*, s_{-x}) + U_\phi^{Q(j)}(s) \geq$$

$$U_{Q(j)l_x}^{Q^*(j)}(s_x^*, s_{-x}) + U_{Q(j)l_x}^{Q^*(j)}(s_x^*, s_{-x}) + U_\phi^{Q(j)}(s) =$$

$$U_\phi^{(Q(j) \cup Q^*(j))l_x}(s_x^*, s_{-x}) + U_{Q(j)l_x}^{Q^*(j)}(s_x^*, s_{-x}) \geq$$

$$U_\phi^{(Q(j) \cup Q^*(j))l_x}(s_x^*, s_{-x}) + U_{Q(j)l_x}^{Q^*(j)}(s_x^*, s_{-x}) \geq$$

$$U_\phi^{(Q(j) \cup Q^*(j))l_x}(s_x^*, s_{-x}) + U_{(Q(j) \cup Q^*(j))l_x}^{Q^*(j)}(s_x^*, s_{-x}) \geq$$

$$U_\phi^{(Q(j) \cup Q^*(j))l_x}(s_x^*, s_{-x}) + U_{(Q(j) \cup Q^*(j))l_x}^{Q^*(j)}(s_x^*, s_{-x}) \geq$$

$$U_\phi^{(Q(j) \cup Q^*(j))}(s_x^*, s_{-x}) \geq U_\phi^{Q^*(j)}(s_x^*, s_{-x}) \geq U_\phi^{Q^*(j)}(s^*),$$

$$\text{则 } \sum_{i \in Q^*(j)} U_{Q(j)l_i}^{Q^*(j)}(s_i^*, s_{-i}) + U_\phi^{Q(j)}(s) - U_\phi^{Q^*(j)}(s^*) \geq 0.$$

$$\text{因此, } \sum_{i \in N} U_i(s_i^*, s_{-i}) + \sum_{i \in N} U_i(s) - \sum_{i \in N} U_i(s^*) \geq 0, \text{ 即}$$

$\sum_{i \in N} U_i(s_i^*, s_{-i}) \geq \sum_{i \in N} U_i(s) - \sum_{i \in N} U_i(s^*)$ 。根据定义 6 可知：

多无人机的自主侦察拥塞博弈是(1,1)-平滑的，则根据定理 1，多无人机的自主侦察拥塞博弈 PoA 至少为 1/2。

现用实例 2 来说明这个界限是紧的。

设有 2 架无人机 A、B，侦察区域共划分为 12 个侦察子区域，其连通性如图 2 所示。无人机 A 的初始位置为子区域 1，无人机 B 的初始位置为子区域 10，无人机 A、B 对于各点的侦察有效概率均为 1。设 $l=5$ ，子区域 1~12 的侦察信息价值分别为：0,0,0,0,0,0,1,0,0,0,1- ϵ ，其中 ϵ 为一个极小的正数。由图 2 可知：由于时间步数的限制，无人机 B 不能在一次行动中同时完成子区域 7 和子区域 12 的侦察。因为子区域 7 的信息价值高于子区域 12，所以无人机 B 的占优策略为：10→9→5→6→7→8，此时间无人机 B 的侦察收益为 1；相应地，无论无人机 A 采取任何策略，侦察收益均为 0。很明显，该局势为该博弈的一个纳什均衡，其社会福利为 1。另一方面，当无人机 B 先侦察子区域 12 时，最多能获得的侦察收益，相应地，无人机 A 能获得的最大侦察收益为 1，此时，该博弈的社会福利达到其最大值 2- ϵ 。由于该博弈只有一个纳什均衡，则 PoA 和 PoS 均为 1/(2- ϵ)，当 $\epsilon \rightarrow 0$ 时，1/(2- ϵ)→1/2；因此，这个界限对于多无人机自主侦察拥塞博弈的 PoA 和 PoS 是紧的。

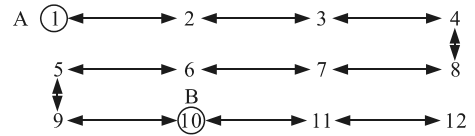


图 2 实例 2

4 策略更新规则

笔者设计了一种多无人机自主侦察拥塞博弈模型中无人机的策略更新规则。该规则是在最优反应动态 (best response dynamics) 的基础上改进得来，笔者将其称为改进最优反应动态 (modified best response dynamics)。

最优反应动态是进化博弈理论中较为典型的动态机制。该机制适用于有快速学习能力的有限理性局中人之间的重复博弈和策略进化。在一次博弈结束之后，局中人会对本次结果进行分析总结，对不同策略的结果做出正确的事后评估并相应调整策略。经过局中人多次不断地调整策略的过程，最终根据“进化稳定策略” (evolutionary stable strategy,

ESS) 给出局中人的均衡解。其中 ESS 的定义如下:

定义 设 s 是博弈 G 的一个局势, 如果存在任意小的 ε , 对于任意的 $s' \neq s$ 和 $\varepsilon \in (0, 1]$, 满足:

$$U(s, (1-\varepsilon)s + \varepsilon \cdot s') > U(s', (1-\varepsilon)s + \varepsilon \cdot s'),$$

则称 s 是一个进化稳定策略。

对于无人机自主侦察拥塞博弈模型, 其典型的可实际应用的解是纯纳什均衡解。由 3.2 节可知: 多无人机自主侦察拥塞博弈中, 可能不存在纯纳什均衡; 为此, 笔者基于最优动态反应, 设计一种易于实现的策略更新规则, 以获取可实际应用的具有合理性能保证的解。

在改进最优反应动态中, 当无人机的最大侦察时间步长为 l 时, 设初始局势为每架无人机通过贪婪的选择, 能使其获取最多侦察收益的长度为 l 的路径而组成的局势, 然后类似于最优反应动态的处理方法, 按照 p_i 的降序来依次计算无人机的最优反应并相应调整策略, 直到达到均衡状态。由于多无人机自主侦察拥塞博弈中的纯纳什均衡不一定存在, 因此, 为避免最优反应动态不能收敛, 笔者约定, 如果不能在 $2Nl$ 轮最优反应动态内找到纯纳什均衡, 则所有无人机将 l 减少到 $l-1$, 然后再试图寻找纯纳什均衡。 l 不断减少, 直到找到纯纳什均衡为止, 此时记为 l' 。接下来, 记 $\tau=l-l'$, 然后再利用最优动态反应在 τ 内寻找纯纳什均衡策略。重复上面的过程, 直到得到长度为 l 的路径为止, 并将其作为各架无人机的侦察路径。在上面的过程中, 若 $l'=0$ 时仍不能找到纯纳什均衡, 则无人机通过贪婪地使自身侦察收益最大化来选择一个子区域作为下一步的策略, 再利用上面的过程, 寻找纯纳什均衡, 直到得到长度为 l 的路径为止。

5 仿真实验

笔者使用随机生成的博弈参数模拟合理的问题场景, 评估改进最优反应动态的性能, 并研究其对各种博弈参数的依赖性。

5.1 参数初始化

随机生成博弈模型, 包括无人机数量、时间步长、子区域的大小和连接性以及侦察信息价值、无人机传感器的有效性参数等。具体如下:

1) 侦察图: 类似于文献[15], 笔者将侦察区域划分为固定维度的网格。

2) 连通性: 通过参数 ψ 来随机产生子区域之间的连通性。对于每一个子区域, 以概率 $\psi=0.5$ 独立

地选择与其 Chebychev 距离为 1 的子区域(行、列、对角线上与其相邻的子区域及其本身)为与其连通的外临子区域。如果遍历所有相邻子区域后仍然没有选择连通子区域, 则需要重复该过程, 直到该子区域至少有一个外临子区域为止, 这样网格中就不会有下沉点, 无人机就不会被困在某个子区域中。

3) 无人机初始位置: 在网格中随机产生无人机初始位置。

4) 传感器有效性: 在每个博弈实例中, 无人机传感器的有效性从(0,1)中随机均匀选择。

5) 子区域侦察信息价值: 每个子区域的侦察信息价值从[0,1]中随机均匀选择。

5.2 仿真结果与分析

通过对比由改进最优反应动态和社会福利最大化 2 种策略更新规则获得的社会福利, 分析最优反应动态规则的有效性。社会福利最大化 (social welfare maximization) 策略更新规则, 即无人机在每一步行动中的策略是长度为 l 的路径, 每架无人机的目的是使得该社会福利最大化。

1) 仿真实验 1。

在实验中, 取 $l=6$, 主要研究 PoA 的实证结果、无人机数量和侦察区域大小对实证 PoA 的影响。

首先记无人机的最大侦察时间步长为 l 时的改进最优反应动态策略更新规则为 M_l , 社会福利最大化策略更新规则为 S_l 。

为了获得 PoA 的实证结果, 笔者利用 5.1 节的参数初始化方法, 随机生成模型数据, 共进行 500 次蒙特卡罗仿真实验, 最终获得有 M_l 和 S_l 2 种规则, 得到的社会福利比率的平均值、最大值和最小值。与此同时, 笔者还比较了不同无人机数量 N 对实证 PoA 的影响, 其中固定网格大小为 7×7 , 分别取 N 为 3,4,5。类似地, 固定无人机数量 $N=4$, 分别取网格大小为 5×5 、 6×6 、 7×7 , 比较了网格大小对于实证 PoA 的影响。实验结果如表 1 所示。

表 1 M_l 和 S_l 所得社会福利的比率

无人机数量 N	Avg	min	max
3	0.974	0.714	1
4	0.951	0.683	1
5	0.928	0.613	1
网格大小	Avg	min	max
5×5	0.923	0.754	1
6×6	0.930	0.792	1
7×7	0.948	0.833	1

从表中可以看出: 实证 PoA 随着无人机数量的减少而增大, 并随着网格的增大而增大。这是因为:

无人机数量较少，或者网格较大时，无人机更可能彼此远离；因此，通过 M_l 得到的各架无人机的侦察策略之间几乎没有重叠的子区域，则这些策略更有可能是各无人机的占优策略，获得的侦察收益更接近各自能够获得的最大侦察收益，故实证 PoA 较大。

从总体上来看，通过 M_l 获得的侦察收益较好，平均实证 PoA 都大于 0.9，而且实验结果中的最小值也远高于 1/2 的理论界限，并且在某些情况下，通过 M_l 也获得了真实最大社会福利。

2) 仿真实验 2。

接下来研究 M_l 获得的社会福利的比率对最大侦察时间步长 l 的依赖性。

笔者考虑较大网格维数情形，将网格大小设置为 10×10 ，无人机数量为 $N=4$ ，并将无人机的最大侦察时间步长 l 分别取 1,3,6,9,12，以研究实证 PoA 对 l 的依赖性。与仿真实验 1 相同，共进行 500 次蒙特卡罗仿真实验，获取 M_l 和 S_l 得到的社会福利比率的平均值、最大值和最小值。实验结果表 2 所示。

表 2 M_l 和 S_l 所得社会福利的比率

最大侦察时间步长 l	Avg	min	max
1	0.998	0.990	1
3	0.979	0.845	1
6	0.951	0.674	1
9	0.928	0.753	1
12	0.912	0.549	1

从表中可以看出： M_l 和 S_l 得到的社会福利比率平均值随着 l 的增大，比率逐渐减小。这是因为：当 l 较长时， M_l 中的 l' 可能远小于 l ，需要多次经过最优动态反应过程才能得到最终结果，从而造成侦察收益的损失。

6 结束语

笔者研究了在通信中断时无人机如何最大化自己的侦察收益的自主侦察任务规划问题，基于最优动态反应设计了一种易于实现的策略更新规则，以获取可实际应用的且具有合理性能保证的解决方案。仿真实验结果表明：针对所研究的问题场景，笔者提出策略更新规则能够取得较好的社会福利，实证 PoA 优于 PoA 的理论边界。

参考文献：

[1] 宗群, 王丹丹, 邵士凯, 等. 多无人机协同编队飞行控制研究现状及发展[J]. 哈尔滨工业大学学报, 2017,

49(3): 1-14.

- [2] 侯岳奇, 梁晓龙, 何吕龙, 等. 未知环境下无人机集群协同区域搜索算法[J]. 北京航空航天大学学报, 2019, 45(2): 124-133.
- [3] 邱斌, 周锐, 丁全心. 多无人机分布式协同异构任务分配[J]. 控制与决策, 2013, 28(2): 274-278.
- [4] ZHU W, LI L I U, TENG L, et al. Multi-UAV reconnaissance task allocation for heterogeneous targets using an opposition-based genetic algorithm with double-chromosome encoding[J]. Chinese Journal of Aeronautics, 2018, 31(2): 339-350.
- [5] 孙海波, 周锐, 邹丽, 等. 通信和测量受限条件下异构多 UAV 分布式协同目标跟踪方法[J]. 航空学报, 2011, 32(2): 299-310.
- [6] SAMAD T, BAY J S, GODBOLE D. Network-Centric Systems for Military Operations in Urban Terrain: The Role of UAVs[J]. Proceedings of the IEEE, 2007, 95(1): 92-107.
- [7] INNOCENTI M, POLLINI L, GIULIETTI F. Management of Communication Failures in Formation Flight[J]. Journal of Aerospace Computing, Information, and Communication, 2004, 1(1): 19-35.
- [8] SUJIT P B, SINHA A, GHOSE D. Multiple UAV task allocation using negotiation[C]//International Joint Conference on Autonomous Agents & Multiagent Systems. 2006: 471-478.
- [9] MIRZAEI M, GORDON B, RABBATH C A, et al. Cooperative Multi-UAV Search Problem with Communication Delay[C]//Aiaa Guidance, Navigation, & Control Conference. 2010.
- [10] BARDHAN R, BERA T, SUNDARAM S. A decentralized game theoretic approach for team formation and task assignment by autonomous unmanned aerial vehicles[C]//2017 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE, 2017: 432-437.
- [11] LI P, DUAN H. A potential game approach to multiple UAV cooperative search and surveillance[J]. Aerospace Science and Technology, 2017, 68: 403-415.
- [12] NISAN N, ROUGHGARDEN T, TARDOS E, et al. Algorithmic Game Theory[M]. Cambridge University Press, 2007: 95-96.
- [13] ROUGHGARDEN T. Intrinsic robustness of the price of anarchy[C]//Proceedings of the forty-first annual ACM symposium on Theory of computing. (ACM), 2009: 513-522.
- [14] MONDERER D, SHAPLEY L S. Potential games[J]. Games and economic behavior, 1996, 14(1): 124-143.
- [15] SCHERER J, RINNER B. Persistent multi-UAV surveillance with energy and communication constraints[C]//2016 IEEE International Conference on Automation Science and Engineering. (CASE), 2016: 1225-1230.