探索問題を対象としたマルチエージェント強化学習法の初期評価

豊味諒磨† 尾崎敦夫‡

大阪工業大学　情報科学部　コンピュータ科学科†（現，情報知能学科‡）

1. はじめに

近年，防災・セキュリティ・清掃等の分野における監視・ 観測・保守・管理業務は，「高危険度」，「長時間」，「広範囲」等，人間には不向きな性質を持つことから，無人システム化が期待されている．それに伴い，複数無人機を協調させ，業務を効率的に遂行するための技術の重要性が増している[1]．

本稿では，マルチエージェントによる領域探索時における最良方策を強化学習により獲得する手法を提案する．

2. 提案手法

2.1 概要

本手法では方策，マップ情報，通信機能を持った自律制御を行うエージェント複数で領域探索を行う．各エージェントの方策の最適化は，方策勾配法定理[2]に基づき行う．

2.2 処理の流れ

2.2.1 領域探索

エージェントは自身が持つ方策をもとに領域探索する．探索中のエージェントが持つ通信範囲に他のエージェントが入った際に互いのマップ情報を共有する．全エージェントの持つマップが探索済みとなった場合に領域探索を終了する．

2.2.2 方策更新

領域探索終了までの総ステップ数T，状態s_iにおけるエージェントnの行動a_jを記録しておき，これをもとに方策を$\theta_{n(s_i,a_j)}$でパラメータ化する．方策勾配法に従い，パラメータ$\theta_{n(s_i,a_j)}$の変化量$\Delta\theta_{n(s_i,a_j)}$を式(1)より求める．

$$\Delta\theta_{n(s_i,a_j)}=(N_n(s_i,a_j)-\pi_n(s_i,a_j)N_n(s_i,a))/\mathrm{T} \quad (1)$$

全エージェントの$\Delta\theta_{n(s_i,a_j)}$が$10^{-4}$以下の場合は学習を終了する．$10^{-4}$以上の場合は$\theta_{n(s_i,a_j)}$を式(2)で更新する．更新量は学習率ηで調整する．その後，再び領域探索を開始する．

$$\theta_{n(s_i,a_j)}=\theta_{n(s_i,a_j)}+\eta\cdot\Delta\theta_{n(s_i,a_j)} \quad (2)$$

3. 評価

提案手法の初期評価として学習の経過回数における各エージェントの$\Delta\theta_{n(s_i,a_j)}$を確認する．また$\Delta\theta_{n(s_i,a_j)}$の変化による領域探索終了までの総ステップ数の変化を確認する

3.1 評価対象

マップ10×10のサイズのグリッドマス，エージェント数は3（各A,B,C）．エージェントの移動方向は上下左右停滞の5つの状態で初期評価を実施した．

図1は，初期配置，視界，通信範囲を示す．

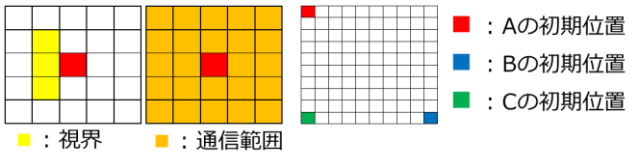

図1. 評価対象

3.2 結果・考察

学習回数と領域探索終了までの総ステップ数，各エージェントの$\Delta\theta_{n(s_i,a_j)}$との関係を図2に示す．図2から全エージェントの$\Delta\theta_{n(s_i,a_j)}$と$\theta_{n(s_i,a_j)}$の差が減少するにつれて総ステップ数が減少していることが確認できた．次に，学習終了後の方策を用い領域探索を行った際の各エージェントの行動結果を以下に示す．各エージェントは初期位置から中央に向け領域探索を行い，中央付近で互いの持つマップ情報を共有し領域探索を終了していた．以上の結果より，全エージェントの$\Delta\theta_{n(s_i,a_j)}$と$\theta_{n(s_i,a_j)}$の差を小さくすることで，マルチエージェントで領域探索を行う方策を獲得できたと考える．

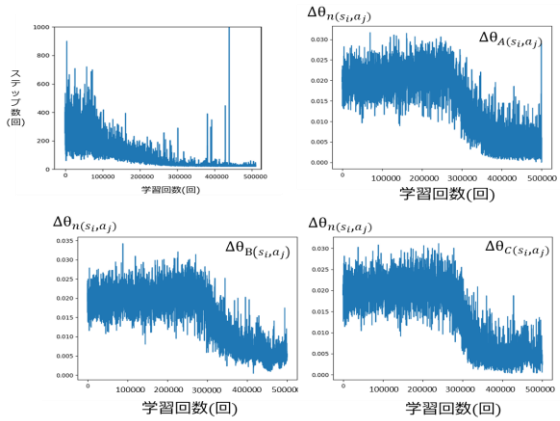

図2. 学習回数に対する総ステップ数と勾配

4. おわりに

本稿では，マルチエージェントによる領域探索時における方策を獲得する手法を提案した．今後の課題は，協調性を確認できるような状況下で学習を行うことと，異なる性質持つエージェント同士の場合で検証を行うことである．

参考文献

[1] Nigam, N., “The multiple unmanned air vehicle persistent surveillance problem: a review”, J. Machines, Vol.2, pp.13-72, 2014.
[2] Sutton, Richard S., et al. "Policy gradient methods for reinforcement learning with function approximation." Advances in neural information processing systems. 2000.