MAS を用いた単体ロボットの行動学習

—反復合議に基づくエージェントの意思決定法の提案 —

室蘭工業大学 情報電子工学系学科 4 年 認知ロボティクス研究室 千葉秀平

1. はじめに

エージェントという制御の主体が複数存在するシステムをマルチエージェントシステム[1]と言う．マルチエージェントシステムでは各エージェントが自律的に行動し，互いに作用することでシステム全体の目的を達成する．マルチエージェントをロボットに適用することにより分散処理による学習速度の高速化や，適応能力の向上などの利点がある．

マルチエージェントシステムを適用したロボットの行動学習手法として強化学習を用いる研究がある．強化学習[2]はエージェントがある状態で取った行動に対して報酬をもらい学習を進める手法である．報酬は目的に合わせて人間が設定する．

マルチエージェントシステムの各エージェントに強化学習を用いることで未知の環境でも自律的にロボットの行動を学習することができる．先行研究に単体ロボット内に複数の強化学習エージェントを設定した研究[3]がある．

2. 先行研究

2.1 先行研究の概要

先行研究は一つのアクチュエータに一つのエージェントを適用することで，状態行動空間を分割する．各エージェントはそれぞれ対応するアクチュエータの動作を学習する．各エージェントが並列して学習することにより学習にかかる時間を短縮できる．

2.2 先行研究の問題点

先行研究では学習の高速化に成功したが，各エージェントが明示的に協調動作をとっていないという問題がある．タスクにおける最適な行動が複数存在する場合，各エージェントの最適な行動も複数存在する．この場合各エージェントの行動が一意に定まらない．そのためロボットの行動として最適な行動とならないことが起こり得る．

この問題を解決するには各エージェントが他エージェントの行動を考慮して協調的に行動選択する必要がある．

3. 研究目的

ロボット内に複数のエージェントが存在するマルチエージェントシステムにおいて，各エージェントが協調して行動を選択する手法を提案する．提案手法により先行研究の問題を解決する．

4. 提案手法

4.1 提案手法のアプローチ

各エージェントの観測する状態に他エージェントの選択した行動を加える．各エージェントが協調して行動選択を行うには，自身の行動に対して他エージェントが選択した行動を知る必要があるためである．

また，各エージェントが複数回行動選択を行い，他エージェントに選択した行動を送信することを考える．行動選択後に他エージェントに選択した行動を送信することで他エージェントの行動を知ることができる．この複数回の行動選択では選択の度に行動を出力しない．複数回の行動選択がすべて終了した後行動を出力する．

4.2 提案手法の概要

提案手法は一つのロボットに対して複数の強化学習エージェントを適用して各エージェントが他エージェントと協調した行動を学習する．本手法では一つのアクチュエータに一つのエージェントを適用する．提案手法の概要を図 1 に示す．本研究ではロボットの一回の行動選択に対し，各エージェントは行動選択を複数回行う．各エージェントは行動選択の都度他エージェントに選択した行動を送信する．この行動選択と選択した行動の通信をまとめて，ステップと定義する．

ステップは既定の N ステップまで行い，N ステップ終了後に行動を出力する．

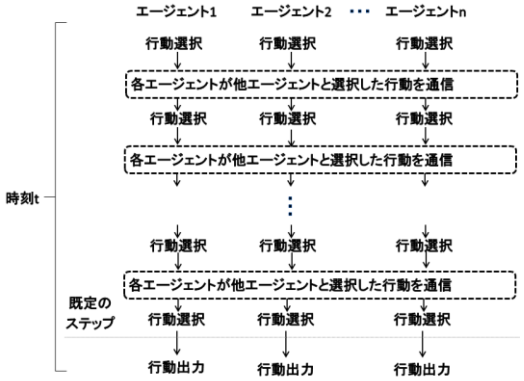

図 1. 提案手法の概要

各ステップの行動選択は本手法で定義する協調的行動評価値$r(a_i)$を用いて行う．協調的行動評価値は強化学習で定義される行動評価値$\mathrm{Q}_i(s, A_i, a_i)$と本手法で定義する行動遷移確率$\pi_{a_i}(A_i)$から算出される．ここで$i$番目のエージェントをエージェント$i$と表わす．環境状態を$s$，エージェント$i$の行動を行動$a_i$，エージェント$i$における他エージェントの行動を$A_i$と表わす．

行動遷移確率は式(1)により 0 ステップの最初に作成さ

れる．Uは行動評価値が最大のものの数である．

$$\pi_{a_i}(A_i) = \begin{cases} \dfrac{1}{U} & (\max(Q)) \\ 0 & (otherwise) \end{cases} \quad (1)$$

協調的行動価値は式(2)により算出される．

$$r(a_i) = \sum_{p \in P_i} (\mathrm{Q}_i(s, p, a_i) \times \pi_{a_i}(p)) \quad (2)$$

行動遷移確率は 1 ステップ以降の各ステップの最初に式(3),(4)を用いて更新する．他エージェントが実際に選択した行動A_iに対しては式(3)を適用し，選択されなかった行動A_iに対しては式(4)を適用する．

$$\pi_{a_i}(A_i) \leftarrow \pi_{a_i}(A_i) + \beta\{1 - \pi_{a_i}(A_i)\} \quad (3)$$

$$\pi_{a_i}(A_i) \leftarrow \pi_{a_i}(A_i) + \beta\{0 - \pi_{a_i}(A_i)\} \quad (4)$$

5. 検証実験

5.1 実験目的

先行研究の手法と提案手法を比較し，複数の最適な行動が存在するタスクにおいて提案手法で協調して行動を選択できていることを示す．

5.2 実験概要

ロボットアームがリーチング動作を行う．アームの先端を目標地点に達することを目的とするタスクを行う．今回の実験では 4 関節のロボットアームを用いる．図 2 に実験環境を示す．各リンクの可動域は全て 0 から 90° である．各リンクはアクチュエータの一回の行動で$-10°$,$0°$,$10°$のどれかに稼動する．エージェントはアクチュエータに設定する．各エージェントは対応するアクチュエータの行動を学習する．

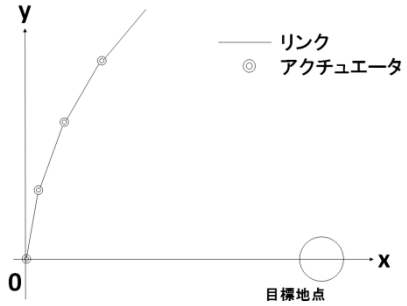

図 2. ロボットアーム

本タスクではアームの先端が目標地点に達することでタスクを達成し，報酬を得る．本タスクの一試行の流れを説明する．

1. ロボットの行動出力
2. タスク達成で一試行終了
3. タスクが達成できなかった場合 1. へ

本実験では各エージェントの学習手法に Q 学習を，行動選択手法には ε -greedy 法を用いる．

本実験での実験パラメータを表 1 に示す．

表 1.実験パラメータ

試行回数	10000
報酬	100
ε	0.05
学習率 α	0.10
割引率 γ	0.90
ステップ数 N	50
β	0.01

5.3 実験結果

各試行における行動数を図 3 に示す．9000 試行から 10000 試行での各試行における行動数を図 4 に示す．図 3 からどちらの手法も学習が進み行動を学習できていることがわかる．図 4 からどちらの手法も本タスク達成の最短行動数である 5 行動でタスクを達成できている．また，提案手法は先行研究の手法と比較して学習が進んでからは行動数が平均的に少なくなっている．

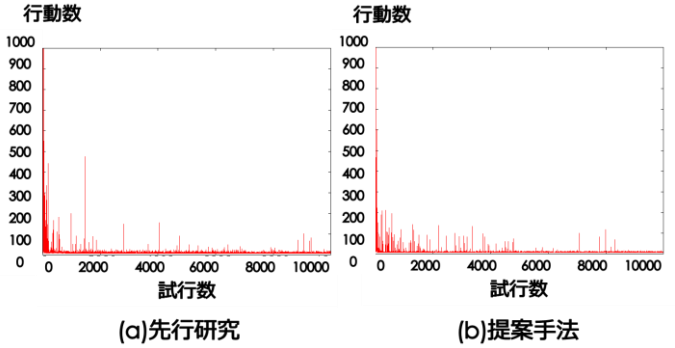

図 3. 各試行における行動数

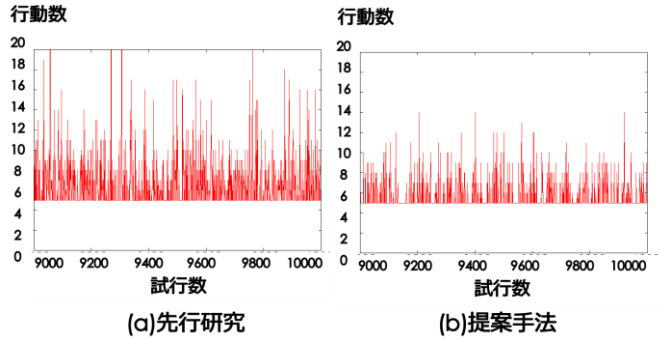

図 4. 各試行における行動数(9000～10000 試行)

5.3 考察

実験結果から先行研究の手法も提案手法もこのタスクにおける最適な行動を学習できている．

また，学習が進んでからの行動数は全体的に提案手法の方が少なくなっている．本タスクでは最短である 5 行動でタスクを達成できる最適な行動が複数ある．このことから提案手法では，複数ある最適な行動を絞り込めていると考えられる．よって，提案手法では他エージェントの行動を考慮して行動を選択するができていると考えられる．

6. まとめ

6.1 全体の考察

本研究では一体のロボットにマルチエージェントシステムを適用し，各エージェントが強化学習を用いて他エージェントと協調して行動を選択できる手法を提案した．実験から提案手法で協調して行動を選択できることが示された．

6.2 今後の課題

今後の課題として以下の点があげられる．

・他の学習手法の適用
・実ロボットへの適用

参考文献

[1] 浅間一：マルチエージェントロボットシステム研究の動向と展望，日本ロボット学会誌，vol10, No4, p428～432, 1992
[2] 畝見達夫:強化学習，人工知能学会誌，1994年
[3] 高泉昇太郎：マルチエージェント強化学習によるシングルロボットの行動学習，卒業論文，2012