マルチエージェントシステムを用いた
単体ロボットの行動学習
−反復合議に基づく協調アルゴリズムの提案−

室蘭工業大学 情報電子工学系専攻 認知ロボティクス研究室 高泉昇太郎

1. はじめに

我々はこれまでにマルチエージェントシステムによるシングルロボットの行動学習手法を提案した[1][1]. (図 1 参照) この手法は単体のロボットに対してロボットに搭載されている各アクチュエータの動作を各エージェントが学習する手法である.

しかし先行研究では各エージェントがエージェント間の協調メカニズムが無いという問題点がある. そのため各エージェントが協調しないと最適な行動が選択できないという場面で最適な行動が選択できないという問題が発生する. 本研究ではこの問題点の解決を目的とする. そのために本研究では単体ロボットに適用するマルチエージェントシステムの各エージェントが協調するアルゴリズムを提案する.

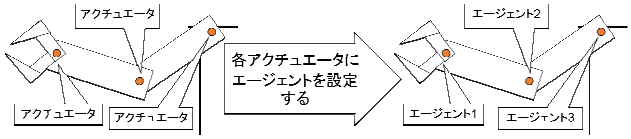


図 1：先行研究の概要

2. 反復合議に基づく協調アルゴリズム

2.1 提案手法の概要

本研究では単体のロボットに対して複数のエージェントを設定する. (図 2 参照) 各エージェントの行動にはそれぞれ各アクチュエータの動作を割り当てる. 各エージェントの状態には環境状態に加えて他エージェントが選択した行動を加える.

ロボットの 1 回の行動出力に対して各エージェントは仮想的に複数回行動選択を行う. この 1 回の行動選択をステップと定義する. 1 ステップ毎に各エージェントは自身の選択した行動を他エージェントに送信する. 各エージェントは送られた行動を基に各エージェントが選択する行動の傾向を求める. この行動の傾向を示す値として行動遷移確率を定義する.

各エージェントは始めのステップで行動遷移確率を算出する. 各エージェントは行動評価値と行動遷移確率から協調的行動評価値を算出する. この協調的行動評価値を元に行動を決定する. 1 ステップ目以降では前ステップで各エージェントが選択した行動を基に行動遷移確率を更新する. 行動評価値と行動遷移確率から協調的行動評価値を算出し行動を選択する. 以降既定のステップ数に達するまでこの手順を繰り返し, 既定のステップ時に各エージェントが選択している行動を実際に出力する行動とする.

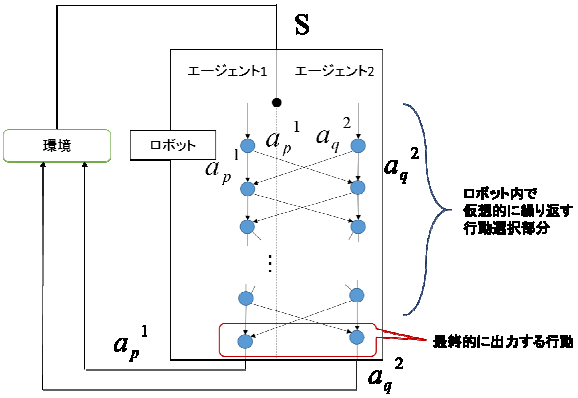


図 2：提案手法の概要

2.2 行動遷移確率と協調的行動評価値の定義

行動遷移確率は他のエージェントが選択する行動の傾向を表す値である. 各行動出力の始めに各エージェントに対して式 (1) によって算出される. またステップ毎に他のエージェントが選択した行動に対しては式 (2), その他の行動に対しては式 (3) を適用し更新する. (MQ は行動価値が最大の行動の数)

$$\pi_{a_c^j}(a_D^k)=\begin{cases}1/MQ(a_c^j\text{の行動価値が最大})\\ 0(\text{その他})\end{cases}\cdots(1)$$

$$\pi_{a^j}(a^k)\leftarrow\pi_{a^j}(a^k)+\beta\{1-\pi_{a^j}(a^k)\}\cdots(2)$$

$$\pi_{a^j}(a^k)\leftarrow\pi_{a^j}(a^k)+\beta\{0-\pi_{a^j}(a^k)\}\cdots(3)$$

また各エージェントが行動を選択する際には協調的行動評価値を算出して使用する. 協調的行動評価値は各ステップで行動を選択する際に算出され, 式 (4) を元に各エージェントの行動評価値と行動遷移確率から算出される.

$$r(a_x{}^j) = \sum_{D=1}^{AC} Q(S_j, a_x{}^j) \times \pi_{a_x{}^j}(a_D{}^k) \cdots (4)$$

3. 実験

3. 1 実験目的

本研究では従来手法と提案手法の比較実験を行い，提案手法が従来手法と同じ行動を獲得し，提案手法の有用性を示す．

3. 2 実験概要

ロボットアームのリーチング動作[1][2]を行う．（図 3 参照）目標地点が試行開始時にランダムに決定され，ロボットアームは各関節を稼働させることで先端部分を目標地点に合わせる．

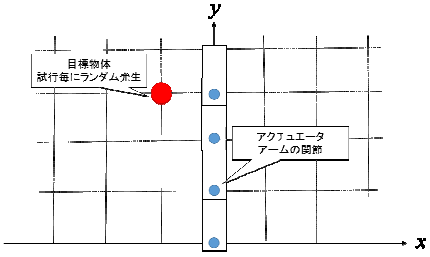


図 3：リーチングタスクの概要

報酬 r は目標地点に先端が到達したときに $r = 100$ を与える．

使用する行動学習手法は加重平均法，行動選択手法は $\varepsilon - greedy$ 法とする．

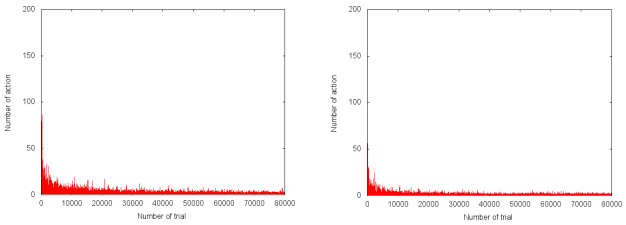
また実験のパラメータを表 1 に示す．

表 1：実験パラメータ

試行回数	80000(回)
タスク達成報酬 r	100
ε	0.05
ステップ数上限 N	100
ステップサイズ・パラメータ α	0.1
β	0.01

3.3 実験結果

各手法の各試行の累計行動数を示す．（図 4 参照）実験設定から 1 試行最低で 1 回の行動出力でタスクを達成することができる．実験結果から両手法共に試行数重ねるごとに収束に向かっていることが分かる．また提案手法は従来手法と比較して試行間の行動数の幅が小さいことが分かる．このことから提案手法のエージェント間の協調行動獲得が有効に働いていると考えられる．



（a）：従来手法　　（b）：提案手法

図 4：各手法の各試行の行動数の推移

また各手法の各試行時点での累計行動数を示す．（図 5 参照）この結果から提案手法は従来手法より累計行動数が下回っていることが分かる．実験設定では 1 つの目標地点に対して複数の最適行動が存在する．このことから提案手法のエージェント間の協調行動獲得によって累計行動数が下回っていると考えられる．

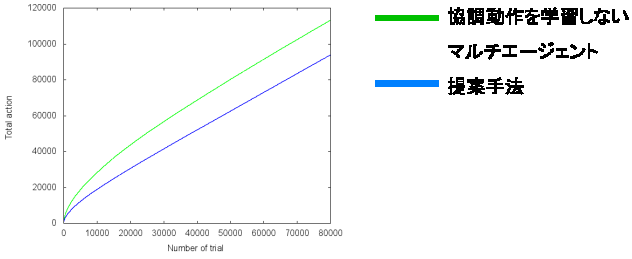


図 5：各手法の各試行時点で累計行動数の推移

3. 4 考察

実験結果から提案手法によってタスク達成に必要となる行動を学習によって獲得できることが示された．

また提案手法は従来手法より試行間の行動数の幅が小さく，また累計行動回数が下回った．これは提案手法のエージェント間の協調行動によって最適な行動が選択されているからだと考えられる．

4. まとめ

4. 1 論文全体の考察

本研究では「エージェント間の協調行動獲得」を目標に，マルチエージェントによるロボット行動獲得方法を提案した．実験結果から提案手法はタスク達成に必要となる行動が獲得できることが示された．

4. 2 今後の課題

本研究の今後の課題として以下の課題があげられる．

（1）他の機械学習への適応
（2）実ロボットへの適応
（3）未知の状態行動対の経験率の向上
（4）各エージェントの探査的行動選択の割合の調整

今後の研究でこれらの問題の解決が望まれる．

参考文献

[1] 高泉昇太郎，倉重健太郎，“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”，日本ロボット学会第 30 回記念学術講演会，RSJ2012AC4F1-6，札幌，北海道，2012.9.17-20

[2] 柴田克成，杉坂政典，伊藤宏司，“強化学習によるリーチング動作の獲得”，電子情報通信学会技術研究報告．NC，ニューロコンピューティング，100(688)，pp107-114，2001