

# MASを用いた単体ロボットの行動学習

## -最適行動のための協調アルゴリズムの提案-

情報電子工学系専攻 認知ロボティクス研究室 12054037 高泉昇太郎

### 背景

**ロボットの行動獲得方法の1つにマルチエージェントシステム(MAS)が存在する。**

マルチエージェントシステム: 複数エージェントによって構成されるシステム。  
MASにおける各エージェントの行動選択に強化学習を使用して複雑なロボットの行動獲得、複数のロボットの協調動作を実現する研究がおこなわれている。

- MASの構成方法
  1. ロボット=エージェント
  2. 一ロボット内マルチエージェント etc.

本研究では一ロボット内に複数エージェントを構成する方法を対象とする。

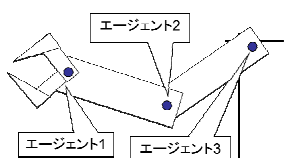
### 問題点

- 従来手法
  - エージェント間で行動の競合や齟齬が起きないように、エージェントの構成方法や行動の決定方法、優先順位が定義されている。
  - 例: 群化マルチエージェント構成による移動ロボット制御
    - 各群の行動決定方法、群間の行動優先順位は手法によって定義されている。
    - 適用するロボットやタスクに応じてエージェント構成方法を設定する。

### 従来研究

- 研究例
  - 強化学習並列化による学習の高速化
  - 群化マルチエージェント構成による移動ロボット制御
  - MAS概念を用いた運動学計算システムを用いた超冗長アームの姿勢列の生成

単体のロボットが認識する状態、出力する行動を複数のエージェントに分割する。



一エージェントが持つ状態、行動の組が少なくなる。

一エージェントのみでは行動獲得が難しいロボットに対して、MASを用いることで行動獲得が容易になる。 アクチュエータごとにエージェントを設定した例

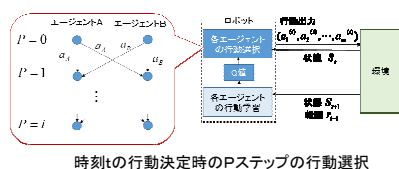
↓

**ロボットがタスクに対して最適なエージェント間の協調動作を自律的に獲得できない。タスクに対して適切な協調動作を設定しなければ最適行動が行えない。**

**MAS手法をロボットに適用したとき、各エージェントがエージェント間の協調動作を自律的に学習して獲得できるようにしたい。**

### 目的達成のアプローチ

1. 時刻tにおける1回の行動選択時、各エージェントが選択した行動情報を共有。
2. 各エージェントの行動情報からの行動選択。
3. 2. を複数ステップ繰り返し出力行動を決定。



時刻tの行動決定時のPステップの行動選択

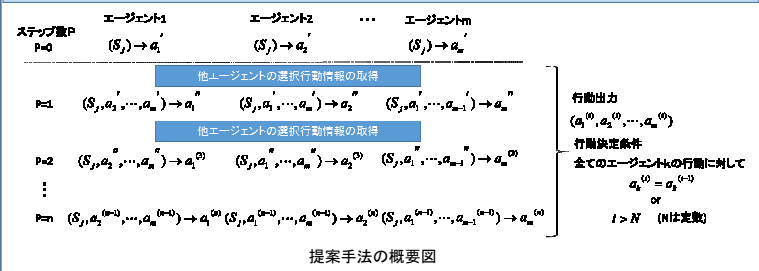
### 研究目的

**MASを用いたシングルロボットの行動学習に対して、各エージェントが他のエージェントが選択した行動を元に行動を決定する方法を提案する。提案手法を適用することでエージェント間の協調動作を自律的に獲得する。**

### 提案手法

#### 【提案手法の概要】

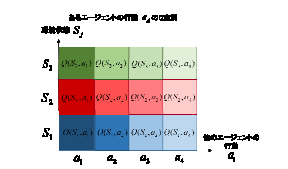
1. 各エージェントが現時点tの環境情報から行動を選択。
2. 全エージェントが行動選択後、各エージェントは他エージェントが選択した行動を取得。
3. 各エージェントは現時点tの環境情報と他エージェントが選択した行動を状態として行動を選択。
4. 2, 3を数ステップ繰り返し行動決定の条件を満たすまで実行。
5. 行動が安定した時点で選択した行動を、現時点tにおける行動として出力。



提案手法の概要図

### 【行動選択-各エージェントの状態行動空間】

- 各エージェントの持つ状態行動対のうち、状態値には、次の2種類の情報によって構成される。
  1. ロボットがセンサ等で得る環境情報
  2. 他エージェントが選択した行動
- 各エージェントはこのQ空間で行動選択を行い、報酬を元に学習する。
- 0ステップ時のみ、行動軸を他のエージェントの行動情報を統合したQ空間を生成し、行動選択する。
- 統合方法はQ値の平均をとる。

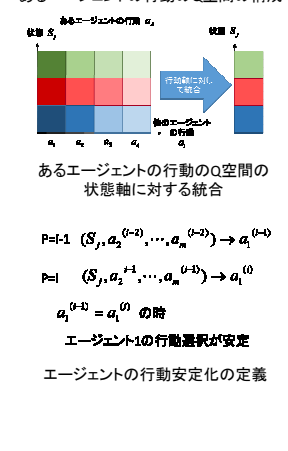


あるエージェントの行動のQ空間の構成

あるエージェントの行動のQ空間の状態軸に対する統合

### 【出力行動決定の条件】

- 以下のいずれかの条件を満たした時、行動を決定する。
  1. 行動安定条件
    - iステップでエージェントが選択した行動と(i-1)ステップで選択した行動が同じとなる。
  2. ステップ回数条件
    - 行動選択時のステップ数iが既定の回数N(Nは定数)に達する。



$P=i-1: (S_t, a_t^{(i-2)}, \dots, a_m^{(i-2)}) \rightarrow a_t^{(i-1)}$

$P=i: (S_t, a_t^{i-1}, \dots, a_m^{(i-1)}) \rightarrow a_t^{(i)}$

$a_t^{(i-1)} = a_t^{(i)}$  の時

エージェント1の行動選択が安定

エージェントの行動安定化の定義

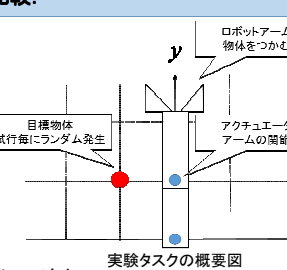
### 実験

#### ◆ 実験目的

提案手法のうち協調動作無し(行動選択時の最大ステップ数N=0)の場合でタスクを実行。従来手法(一ロボット=エージェント)と実験結果を比較。

#### ◆ 実験タスク

ロボットアームの物体回収タスク。ロボットの目的はランダムに発生する目標物体を回収すること。ロボットアームには2つの関節が存在、アクチュエータにより動作する。



実験タスクの概要図

#### ◆ 実験概要

- 一ロボット=エージェント、一ロボットマルチエージェント(提案手法)でタスクを実行。
- マルチエージェントのエージェントの数はアクチュエータの数だけ設定。
- 各手法の実験結果を比較。

#### ◆ アクチュエータの動作

- 行動: 0°(右) 90°(正面) 180°(左)の方向へ可動

#### ◆ ロボットが認識する環境状態設定

- 各アクチュエータの状態(3状態×2関節)
- 目標物体のx座標、y座標

#### ◆ 報酬設定

- タスクを達成した際の報酬  $r=100$

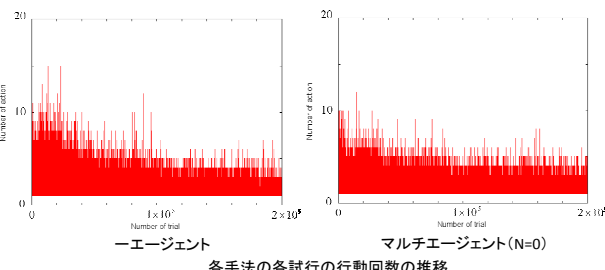
#### ◆ 1試行の流れ

1. 試行の初めに位置がランダムに目標物体を生成。ただしアームが届く範囲とする。
2. エージェントが行動選択後、各アクチュエータを稼働。これを1行動とする。
3. タスクを達成していない際はその地点から再び行動選択を行い行動する。
4. アームの先端の座標が目標物体と重なった時点でタスク達成とする。
5. タスク達成後、アームの状態はそのまま1に戻り次の試行を始める。

### 実験設定

試行回数	200000(回)
1試行の行動回数	物をつかむまで
行動の間隔	全アクチュエータの行動後
学習タイミング	1行動毎
目標物体	アームが届く位置に1試行毎にランダム
最大ステップ数N	0
初期値	実験開始時のアーム 1試行終了ごとのアーム
タスク達成報酬	$r=100$
$\delta$	0.05
ステップサイズ・パラメータ	0.1
割引値	0.9

#### ◆ 実験結果



各手法の各試行の行動回数の推移

### 考察

- マルチエージェントを適用した場合は、各試行の行動数の推移が一エージェントの場合とほぼ同等の行動数となった。
- 行動協調なしの時、マルチエージェント手法は一エージェントの手法と同等の結果が得られる。

### 今後の課題

- 協調動作有り(N>0)の場合で実験を行う。
- 一エージェント、N=0の場合と実験結果を比較する。

~実験完了後~

- 関節数(エージェント数)を増やしたロボットで実験を行う。
- 各エージェントが自身に必要な状態を選択して取得できるアルゴリズムを提案する。