# マルチエージェントによるシングルロボットの行動学習に関する研究

室蘭工業大学 情報工学科 4年 認知ロボティクス研究室 高泉昇太郎

## 1. はじめに

強化学習とは未知なる環境における適切な行動戦略を，経験を繰り返すことで獲得する学習アルゴリズムである[1][2]．（図1参照）．強化学習はロボットの制御方法として注目されている．

強化学習の問題点の1つに，「状態値，行動数の増加による学習時間の増加」がある．本研究ではこの問題を解決することを目標とする．本研究ではマルチエージェント法を利用する．複数エージェントが学習し，学習領域を分割することで学習効率の向上を狙う．マルチエージェント法を導入するために，本研究ではロボットのアクチュエータ毎にエージェントを設定する手法を提案する．

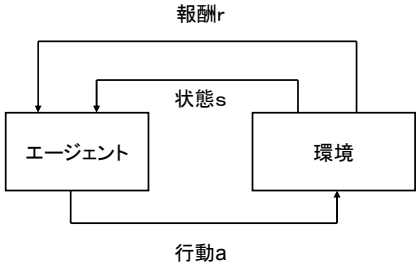



**図 1：強化学習の概要**

## 2. 提案手法

### 2.1 ロボットの行動

「ロボットの行動」とは，搭載されているアクチュエータを動かすことである．ロボットの行動はアクチュエータの数や種類で変化する．また複数のアクチュエータの同時稼動による相互作用により1行動を構成する．（図2参照）

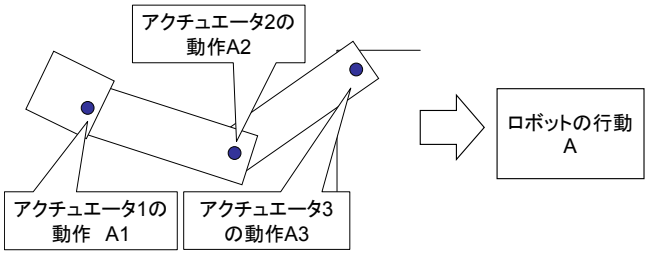



**図 2：各アクチュエータの相互作用**

### 2.2 提案手法の概要

本研究では各アクチュエータにエージェントを設定する手法を提案する．アクチュエータ毎に強化学習を行い最適な動作を獲得してロボットのタスク達成を可能にする．各アクチュエータにエージェントを設定することで状態行動対をアクチュエータの動作毎に分割し，同時に学習を行わせる．（図3参照）

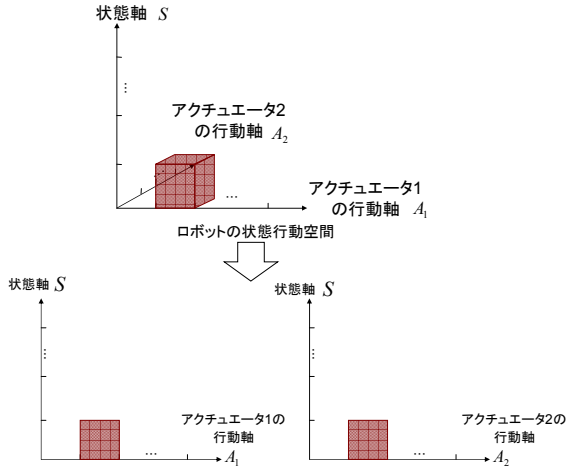



**図 3：状態行動対の分割**

### 2.3 提案手法の概要

提案手法では取得した状態を元に各エージェントが動作を決定する．動作を決定した後実際に行動する．その後環境から報酬を受け取り，各エージェントが実行した動作に対して学習を行う．（図4参照）

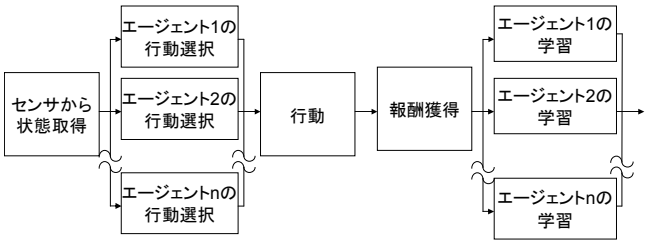



**図 4：提案手法の概要**

各エージェントが相互作用し学習を行うために，各エージェントに全エージェントの状態を認識させ，全アクチュエータの動作は同期をとり一斉に処理を行わせる．

## 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，学習の収束が早くなっていることを検証する．

3.2 実験概要

台車ロボットの荷物運搬タスクを行う．（図５参照）台車には角度を変えられるテーブルが搭載されている．台車は加速度することで目的地に到達する．その際にテーブルの上のものが落ちないようにテーブルの角度を調節する．

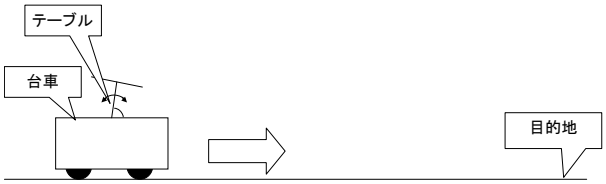



**図 5：荷物運搬タスクの概要**

報酬 $r$ は台車の現在位置のセンサ値を $x$，目的地の位置センサ値を $G$，テーブルの角度を $\theta$，加速度と重力の合力を $R$ とした時，式（1）で決定される。$w_1$，$w_2$，$w_3$ は係数である．

$$r = w_1(\theta - R)^2 + w_2(G - x)^2 + w_3 \quad \cdots(1)$$

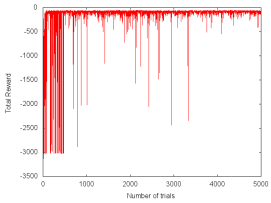
また実験のパラメータを表 1 に示す．

**表 1：実験パラメータ**

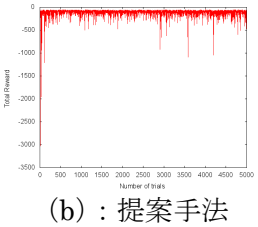
| | |
|---|---|
| 試行回数 | 5000（回） |
| 1 試行の行動回数 | 1000（回） |
| 行動の間隔 | 1 秒毎 |
| 学習タイミング | 1 行動毎 |
| 目的地の位置 | 状態値 $x_G = 5.5$ |
| 重力加速度 $g$ | $9.8(m/s^2)$ |
| 初期値 | 加速度 $a$ : $0(m/s^2)$<br>角度 $\theta$ : $90(°)$<br>速度 $v$ : $0(m/s)$<br>位置のセンサ値 $x$ : 0<br>Q 値： 0 |
| 学習手法 | Q-learning |
| 行動選択手法 | $\varepsilon$ -greedy 法 |
| 係数 | $w_1 : -10 \quad w_2 : -0.1 \quad w_3 : 0$ |
| $\varepsilon$ | 0.05 |
| ステップ・サイズ・パラメータ $\alpha$ | 0.5 |
| 割引値 $\gamma$ | 0.5 |

3.3 実験結果

各試行の 1 試行毎の総獲得報酬の推移示す．（図６参照）この結果から獲得する報酬が安定し始める試行数は提案手法のほうが従来手法より速いことが分かる．また未探索の状態行動対になることで獲得報酬が少なくなる試行がどちらの手法でも存在するが，提案手法のほうが従来手法よりも回数が少なく，また提案手法のほうが従来手法より報酬が低くなっていないことが分かる．この結果から提案手法では強化学習が正しく行われ，従来手法より早く学習が収束することが示された．




（a）：従来手法




（b）：提案手法

**図 6：各手法の 1 試行毎の総獲得報酬の推移**

3.4 考察

実験結果から提案手法は従来手法と同じ行動を獲得し，試行数を削減していることが示された．これはマルチエージェントの協調動作が正常に行われ学習が正しく行われていることを示している．またマルチエージェントによる状態行動対の切り分けが正しく行われ，学習が収束するまでの時間を削減できていることを示す．一方で提案手法は従来手法より学習精度が劣るという結果も示された．提案手法では各エージェントが $\varepsilon$ の確率でランダムに行動するため，システム全体がランダムに行動する確率が従来手法より大きくなっているためである．

4. まとめ

4.1 論文全体の考察

本研究では「強化学習の学習時間の短縮」を目標に，マルチエージェントによるロボット制御方法を提案した．実験結果から提案手法は学習が収束し従来手法より収束が早いシステムだと証明できた．一方で提案手法は従来手法より学習精度が劣る結果も示された．以上の結果から提案手法は学習精度より学習速度が重要なタスクに適しているシステムといえる．

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）学習精度の低下

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] 畝見達夫，“強化学習法とロボットへの応用”, 日本ロボット学会誌, Vol.13, No.1, pp51-56, 1995

[2] 森紘一郎，山名早人，“強化学習並列化による学習の高速化”, 情報処理学会研究報告. ICS, [知能と複雑系]，pp89-94，2004