

群の中の個体の知能の発達

木島康隆*

1 はじめに

ハードウェア技術の発展と共に、工場だけではなく、家庭やオフィス、自然環境とロボットが直面する環境も多様・複雑化してきた。そのため、多様・複雑な環境に対して適した行動をとるためにさまざまな研究が行われている。

本研究では群を利用した個体の知能発達に着目する。個体の知能の発達について考えると、知能は学習することによって発達する。学習は自身が得た情報をもとにして行われ、得られる情報量に比例して最適行動の探索がしやすくなり、学習の効率が上がる。

個体単体で学習を行う場合は、自身の試行錯誤によって得られる情報からしか最適行動の探索を行うことができない。一方、コミュニケーションを利用した学習では、コミュニケーションにより他者から情報を取り入れることで、自身が試行錯誤によって得る情報よりも多くの情報を獲得することが出来る。そして、それらの情報をもとに最適行動の探索を行うことが出来、単体で学習するよりも学習の効率が上がる。

このようにコミュニケーションを用いることで個体の学習が促進される。そこで、本研究では、コミュニケーションを用いた個体知能の発達を促進させるシステムの構築を目的とする。本研究では、環境に適応し発展するために、個体が群としての特性を利用することを考える。そのため、群でなくても機能し、群となることでより効率的に学習を行える知能システムの構築を目指す。

2 本論文で扱うコミュニケーション

さまざまなコミュニケーションの仕方があるが、本論文では一例として、出来るだけ簡単なコミュニケーションの仕方について考える。まず、コミュニケーションするための条件として以下を考えた。

- 個体間で扱う情報の形式が同一である

そして、簡単なコミュニケーションの条件として以下を考えた。

- 情報の処理の仕方が個体間で共通であること

そしてこれらの条件を満たす個体の十分条件としては、以下の2つを考える。

- 身体構造が同一
- 目的・タスクの種類が同一

この条件のもとコミュニケーションする情報を考える。状況・身体に強く依存する情報は、個体固有のものであり、コミュニケーションには不向きである。よって、状況・身体への依存度の低い情報を考える。そのような情報の中から本研究では、「学習法」に注目する。

3 提案システム

学習法をコミュニケーション情報とした学習システムの概要図を図1に示す。学習システムは、行動学習部と学習法学習部の2つの部分で構成している。行動学習部は、直面する状況に対する学習であり、過去の行動結果から状況に応じた行動を学習する。学習法学習部では、直面する環境に適した学習法を学習する。学習法の学習は、行動学習部で用いた学習法に対して、そのとき得られた結果を使い行う。更にコミュニケーションにより、状況・使用した学習法・その時得られた結果に関して情報を交換することで、他者の情報も自己の知識に反映させる。本研究では、この学習システムに強化学習を適用する(図2)。強化学習適用システムでは、学習法は行動選択手法と行動評価手法の組となる。また、結果は報酬となる。

4 提案システムの有効性の検証

強化学習を適用したシステムの有効性を検証するために、非定常N本腕バンディット問題を実験対象とした実験を行った。腕の当選確率が変動振幅(Amp)、変動頻度(Th)の2つの値によって変化する。変動振幅は腕の当選確率変動の大きさを表し、変動頻度は腕の当選確率の変わりやすさを表す。本実験では、変動振幅(Amp)、変動頻度(Th)による環境マップを作成し、そこに複数台のエージェントを配置する(図3)。エージェントは環境マップの各位置に応じた状況について学習を行う。学習法学習部では10手法(表1)の中からエージェントが直面する環境にあった学習法を選択する。

*室蘭工業大学 情報工学科 認知ロボティクス研究室

実験結果としては、まず、学習法の選択の推移 (図 4, 図中のカラーバー横の数字は表 1 の学習法番号に対応する) を観察すると、試行 15000 回以降は選択している手法の分布が安定したため、エージェントがそれぞれ直面する環境に合った学習法を学習していることが分かった。次に、コミュニケーションのなしの場合 (図 5) と人間が学習法を設定した場合の 2 つの場合と平均獲得報酬量で比較した結果、全体的に提案システムの方がプラスの結果が出た。これら 2 つのことより、提案したシステムの有効性を検証することが出来た。

5 今後の課題

今後の課題としては、実ロボットを用いたタスクでの実験を行うことを考えている。また、今回は出来るだけ簡単なコミュニケーションとして、同タスク・同目的という条件であったが、他タスク・他目的の個体同士でもコミュニケーションできるような、より高度なコミュニケーションについて考察していきたい。

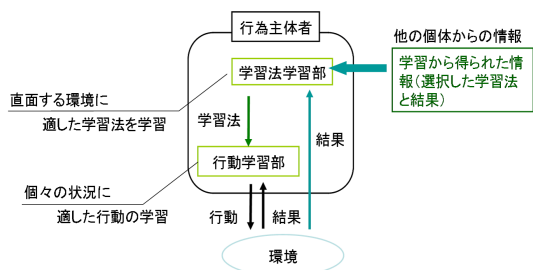


図 1: 学習法をコミュニケーション情報とした学習システム

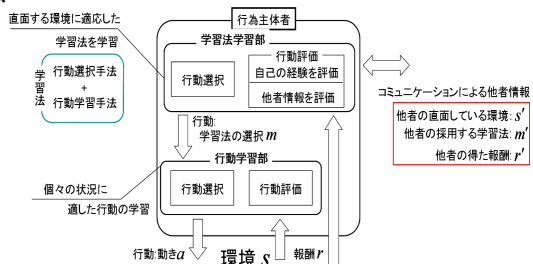


図 2: 強化学習適用システム

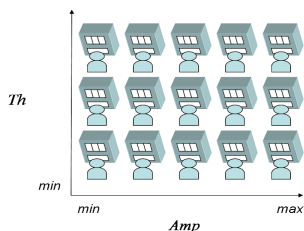


図 3: 実験環境

表 1: 学習法学習部で選択する学習法

学習法番号	行動選択手法	行動評価手法
0	softmax 法	標本平均手法
1	softmax 法	加重平均手法
2	softmax 法	Q 学習法
3	greedy 法	標本平均手法
4	greedy 法	加重平均手法
5	greedy 法	Q 学習法
6	追跡手法	標本平均手法
7	追跡手法	加重平均手法
8	追跡手法	Q 学習法
9	強化比較法	なし

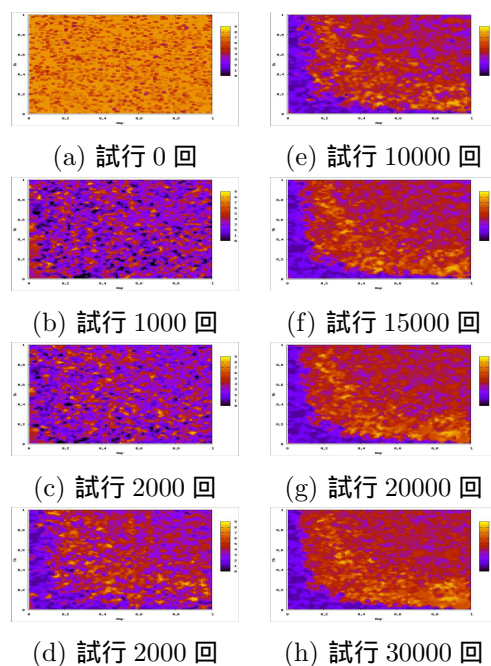


図 4: 学習法選択の推移

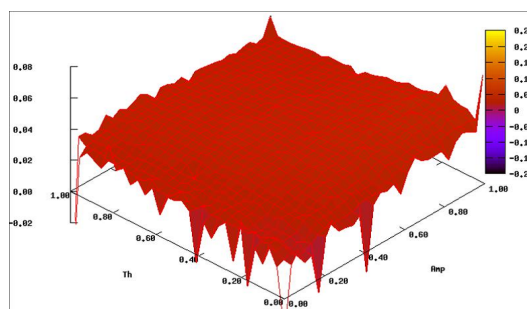


図 5: コミュニケーションなしの場合との平均獲得報酬の比較