

進化計算に基づくロボットの様々な行動学習に対する取り組みと考察

理学専攻 情報科学コース
曾我 紗知子

1 はじめに

進化計算を用いて環境に適応するロボットコントローラを得る際に、獲得されたコントローラは学習を行った環境での振る舞いは保証されるものの、異なる環境に置かれた際には適切な振る舞いは保証されない。また、実世界においてロボットを動作させる際は、単純な目的のみではなく、複数の目的を同時に達成する行動が必要となる場合が多い。その中で、特定の行動学習をある固定された適応度関数を用いて行う場合、学習する環境によって、その学習効率が変わることが想定される。つまり、効率的な学習を実現するためには学習環境を工夫する必要があると言える。

以上のことから本研究では、単純な目的の下、学習環境とは異なる環境でも動作可能であるロボットコントローラの獲得を踏まえ、複数の目的を持つ行動学習を考察する。さらに、環境設定に対するロボットの行動学習の効率に関する評価、及び検討を行う。

2 ロボットの行動学習

2.1 学習環境

本研究では、Web 上でオープンソースとなっている、進化ロボティクスの研究・教育用 3D シミュレータ Simbad[1] を利用して、ロボットコントローラの学習を行う。今回のシミュレーション実験では、センサ(ソナー)と速度、回転速度を利用する。

2.2 コントローラの構成

ロボットのコントローラは、図 1 に示すニューラルネットワークで構成され、遺伝的アルゴリズムによって進化させる。

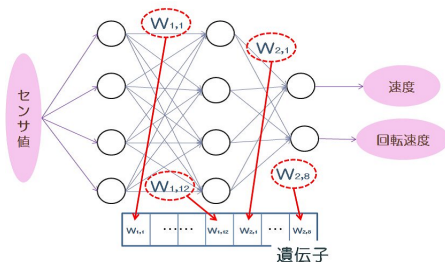


図 1: ロボットコントローラ

コントローラは、単一方向にのみ信号が送られるフィードフォワード型の入力層 4、中間層 4、出力層 2 のニューラルネットワークで構成する。ロボットが取得するセンサ値を入力とし、速度と回転速度を出力とする。遺伝的アルゴリズムを利用する際に必要となる遺伝子には、入力層、中間層、出力層全てのノード間の 24 個の結合荷重の値を遺伝子座に入れる。

2.3 学習アルゴリズム

本研究では、単目的の行動と多目的の行動について学習を行った。単目的の行動学習では、学習アルゴリズムに単純遺伝的アルゴリズム(単純 GA) を利用し、多目的では、多目的遺伝的アルゴリズム(多目的 GA) の NSGA-II を利用した。

3 行動学習実験

単目的の行動学習の詳細な説明は、文献 [2] に譲り、ここでは多目的の行動学習について説明する。

3.1 実験設定

ここでは、「壁とロボットに衝突しないこと」と「目的地に行くこと」、この 2 つを目的とし、学習を行う。学習環境は図 2 で行い、目的地は右上のロボットが置かれている座標 $(2, -2)$ とする。ロボットは全部で 5 台用意するが、そのうちの 4 台は、障害物回避を学習済みのコントローラで動作させる。白い矢印で示されている 1 台のロボットのみを学習させる。

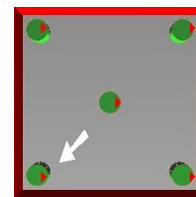


図 2: 学習環境

使用した NSGA-II の設定を以下に示す。

個体数: 60 交叉率: 0.9

突然変異率: 0.1 終了世代数: 200

適応度関数の設定は以下の通りである。

$$f_1 = \sum \{v + (\pi - |r|) + s\} \quad (1)$$

$$f_2 = \sum_t (d_t - d_{t-1}) \quad (2)$$

f_1, f_2 : 適応度 v : 速度

$|r|$: 回転速度の絶対値 s : センサ値の平均

d_t : t ステップ時の $(2, -2)$ からの距離

式 (1) は、方向転換が少なく、壁に衝突しない場合に適応度が高くなるものであり、式 (2) は、1 ステップ毎に目的地に近づくほど適応度が高くなるように設定したものである。計算の終了条件は、1000 ステップまでに計算が終了するか、ロボットの速度が 0、1 ステップ間の移動距離が 0、または目的地に到達することである。また、壁やロボット同士で衝突した際には、どちらの関数も適応度を 0 にし、終了条件が満たされるまで計算を続けるものとする。

3.2 実験結果

障害物となる 4 台のロボットの初期位置を固定した環境下で 200 世代まで学習を終えたコントローラは、障害物となるロボットを避けながら目的地へ到達し、上述した 2 つの目的を達成できたことが確認された。学習後に獲得されたコントローラを使用し、図 3 の環境で動作確認をした。ここでは、左下の $(-2, 2)$ のロボットに学習後のコントローラを適用し、初期位置はランダムに出現させた。図 3 に学習後のロボットの動作を示す。この結果より、学習環境とは違う環境(ロボットの初期位置が異なる)でも動作可能であることがわかった。

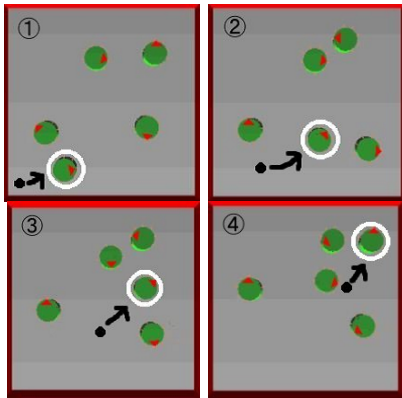


図 3: 学習後の動作

4 行動環境における学習効率

特定の行動学習を固定された適応度関数を用いて行う場合、学習の環境によってその学習効率は変わる。単純な行動の学習ほど適応度関数は同じものを採用する場合が多くなり、このことからも効率的な学習を実現するためには学習環境を工夫する必要があるため、学習効率について検討、考察を行った。

4.1 静的な環境下での学習

時間によってロボットを取り巻く環境が変化しない、静的な訓練環境 (図 4) とその環境を基に、直進を必要とする要素と道幅が狭い要素を加えて訓練環境より複雑にした環境 (図 5) の 2 つを用意する。

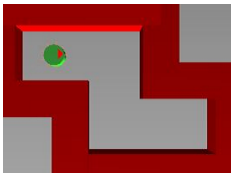


図 4: 静的訓練環境

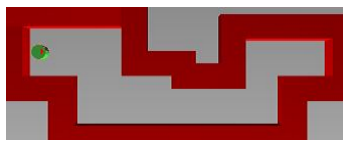


図 5: 静的環境

ここでの適応度関数は、文献 [2] で衝突回避を学習可能であることが判明している式を使用した。

訓練環境で 20 世代まで学習を行うと、壁を避けて環境内を巡回するコントローラを獲得できた。また、図 5 の環境でも学習を行うと、適応度関数が一番高くなるまでの世代数が 120 世代となった。次に、訓練環境で得られたコントローラを使って、ロボットを図 5 の環境で動かしてみると、こちらの環境でもロボットは壁に衝突することなく、滑らかに動き続けた。このことから、「壁への衝突を避けて移動する」学習をするためには、その行動が実現されるのに十分な環境が与えられていれば良いことがわかる。つまり、学習目的とする行動が単純な場合は、ロボットは複雑な環境で学習を行うよりも、その複雑な環境を想定して単純な環境を構築し学習を行う方が効率的であると考えられる。

4.2 動的な環境下での学習

ここでは、学習を行うロボットの他に、複数台のロボットをランダムに動作させ続けることにより、ロボットを取り巻く環境が変化する動的な環境を構築した。また、移動物体との衝突を回避することに重点をおき、文献 [2] の式から総移動距離を除いた適応度関数を設定した。

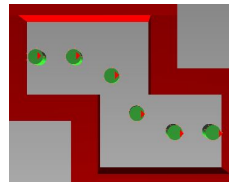


図 6: 動的訓練環境

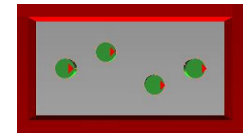


図 7: 動的環境

まず初めに、訓練環境を図 6 とし、シミュレーションを行う。6 台で学習を行った結果、30 世代目で、壁に衝突しないことに加えて、ロボット同士も衝突しないコントローラを得ることができた。このコントローラでは、ロボットはある程度環境内を巡回し、衝突しそうになると避ける動きを見せる。台数を少なくすると、衝突回避の学習は行うことができなかった。これは、同じ環境において、ロボットが少ない台数の時には、図 6 の訓練環境が広く、学習の機会が十分に与えられなかったため、学習が不十分になってしまい、適切に動作するコントローラが得られなかったということが考察される。

次に、図 6 の環境よりも約半分の大きさの図 7 の環境で学習を行った。4 台で学習を行うと、37 世代で適応度が一番高いコントローラを獲得することができ、ある半径で回転して、他のロボットと衝突しそうな場合には避ける、という学習結果となった。他の台数では、満足する動作を行うコントローラを得ることができなかった。このことは、学習に適したロボットの台数があるということを示しており、学習に対する適切な環境の必要性を明示している。

図 6 の環境で 6 台のロボットで学習を行ったコントローラをそのまま利用し、図 7 の環境で 4 台にして動作確認をしたところ、正しく動作した。動的な環境でも、うまく学習が行えなかった場合、動作させたい環境を基にし、訓練環境を構築して学習を行うことで、動作可能にすることができるということがわかった。

5 おわりに

本研究では、ロボットのコントローラをニューラルネットワークにより構成し、遺伝的アルゴリズムで学習を行った。それにより、目的を達成するコントローラを獲得し、訓練環境だけでなく、他の環境へと移しても動作可能であることを確認した。また、目的が 1 つのものだけではなく複数になった時にも、多目的遺伝的アルゴリズムを利用することにより、ロボットを動作させることができた。そして、訓練環境が学習する間に変化をせず、目的が単純な場合は、複雑な環境を想定した単純な環境で学習を行うと良いことがわかった。さらに、訓練環境が学習する間に変化をする場合は、十分な学習の機会が必要となり、環境と移動物体の数の関係を工夫することで効率の良い学習が実現できることを確認した。

参考文献

- [1] <http://simbad.sourceforge.net/>
- [2] 曾我 紗知子, 小林 一郎, 様々な環境下におけるロボットの行動学習に関する取り組み, 第 25 回人工知能学会全国大会, 2I3-2, 2011