2T-01

# 動作中に構造が変化する進化型ニューラルネットワークによる自律エージェントの行動獲得

広井香菜子†　　長尾智晴†
†横浜国立大学　大学院環境情報学府

## 1　まえがき

近年，自律エージェントの行動獲得に関する研究に注目が集まっている．自律エージェントの行動獲得に用いられる手法の 1 つとして進化型ニューラルネットワークが挙げられる．進化型ニューラルネットワークとは，進化計算を用いてネットワークの構造や結合荷重を獲得する手法である．筆者らの研究グループでは，進化型ニューラルネットワークの 1 つとして Real valued Flexibly Connected Neural Network（RFCN）[1] の提案を行っている．RFCN は連続値空間におけるエージェント制御問題に適用され，有効性が示されている．しかし RFCN には一度学習で得られた構造を動作中に変更する仕組みはない．そのため獲得した構造で対応できない環境におかれた場合，再学習を行う必要がある．本稿では，動的に新たな環境に適応することを目指し，ネットワークの構造をエージェントの動作中に変化させる手法の提案を行う．実験では障害物回避問題に適用し，提案手法の性能の検証を行った．

## 2　動作中に構造が変化する進化型ニューラルネットワーク

提案手法では，RFCN に構造を動作中に変更する仕組みを取り入れた．図 1 に提案手法の構造を，図 2 にエージェントの動作中に行う構造変更の流れを示す．ネットワークの結合には動作中に変更しない固定の結合と，変更する結合の 2 種類がある．結合の種類を表す遺伝子を RFCN に組み込み，構造と同時に最適化を行う．

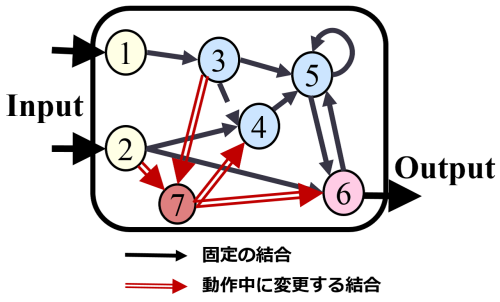



図 1: 提案手法のネットワーク構造

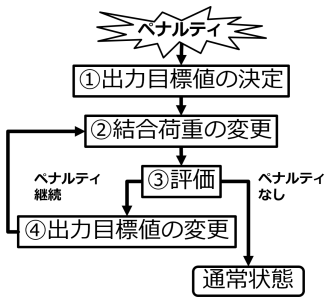



図 2: 動作中の構造変更の流れ

次に，動作中に行う構造の変更方法について説明する．提案手法はエージェントの行動中に，環境から与えられるペナルティをもとに構造の変更を行う．ペナルティが与えられた場合，センサ情報と直前の行動ログを用いて出力目標値を決定する．そして決定した目標値に近づくように動作中に変更するユニット間の結合荷重を更新する．次に，変更後の構造でエージェントを動かし，変更の評価を行う．ペナルティが継続して与えられる場合，出力目標値を変更し，その値をもとに再度結合荷重の変更を行う．これをペナルティが与えられなくなるまで繰り返す．結合荷重が 0 の場合，ユニット間の結合が無いことと同じであるから，この変更でネットワークの構造を変化させている．

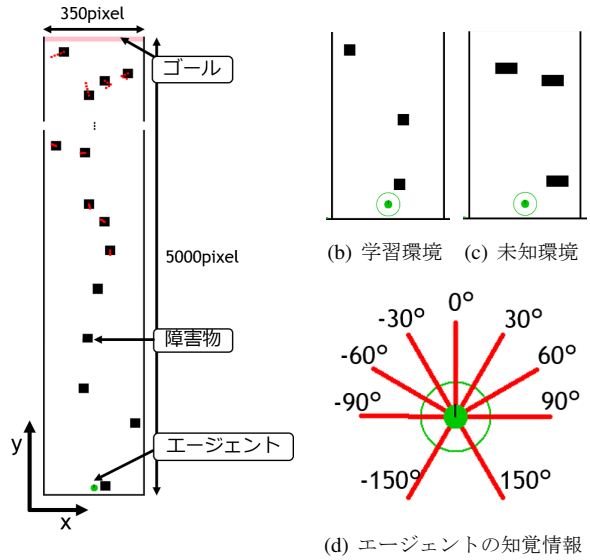
## 3　障害物回避問題におけるエージェントの行動獲得実験

### 実験設定

実験では，シミュレーション環境上で障害物回避問題を扱った．図 3 に実験環境を示す．エージェントは障害物を避けながらマップ上方のゴールを目指す．ゴールに到達した場合はタスク達成とし，障害物や壁に衝突した場合はそこでタスク失敗とした．図 3(a) に実験環境のマップ全体を示す．図 3(b) にエージェントが学習に用いたマップを，図 3(c) には未知環境として用いたマップの例を示した．未知環境は学習に用いる環境と比べ障害物を大きめに設定した．エージェントは入力情報として，図 3(d) に示した 9 方向の距離センサから得られる障害物までの距離と行動に対するペナルティが与えられる．ペナルティは，次の 2 つの場合に与えた．

- 一定範囲内に障害物が存在する場合
- ゴールから遠ざかっている場合

Action Control of Autonomous Agents using Evolutionary Neural Network Adapted Topology during Running.
†Kanako Hiroi　†Tomoharu Nagao
†Graduate School of Environment and Information Sciences, Yokohama National University




(a) 全体図　(b) 学習環境　(c) 未知環境　(d) エージェントの知覚情報

図 3: 実験環境

距離センサが測定できる最大距離は 100pixel，障害物接近のペナルティが与えられる範囲は 25pixel 以内とした．比較手法として RFCN を用いた．また，提案手法は次の 2 種類の学習方法で比較を行った．

- 提案手法 1
  変更する結合を含めたネットワークを GA で獲得
- 提案手法 2
  提案手法 1 で獲得した個体を未知環境に $N$ 回適用し，構造の変更を行う

今回は $N = 5$ とした．提案手法 2 は提案手法 1 と比べ構造を変更する回数が多いことから提案手法 1 よりよくなることを期待している．学習環境として基本の固定環境と，それにランダムな変動を加えた類似環境 2 つの合計 3 種類のマップを用いた．エージェントはゴールにどれだけ近づいたかで評価される．用いた適応度関数を式（1）に示す．

$$Fitness = \frac{1}{M}\sum_{i}^{M}(l_i + R_i - \alpha L_i) \quad (1)$$

ここで，$M$ は学習に用いるマップの数，$l_i$ は $i$ 番目のマップでエージェントが y 軸方向に移動した距離，$L_i$ は総移動距離である．$\alpha$ は 1 より小さい正の定数である．エージェントが無駄な動きをせずにゴールに向かうために，総移動距離 $L_i$ によって適応度の減点をした．$R_i$ はエージェントがゴールした場合に与えられる報酬である．今回の実験では $R_i = 1000$ とした．

### 実験結果

表 1 に，学習で得られた最良個体を未知環境 1000 マップに適用した結果を示す．未知環境でも学習と同じく式（1）を用いて評価した．図 4，5 に最良個体の

表 1: 実験結果 (10 試行)

| | 提案手法 1 | 提案手法 2 | RFCN |
|---|---|---|---|
| 平均 | 3163 | 3183 | 2796 |
| 最良個体 | 3545 | 3681 | 3485 |

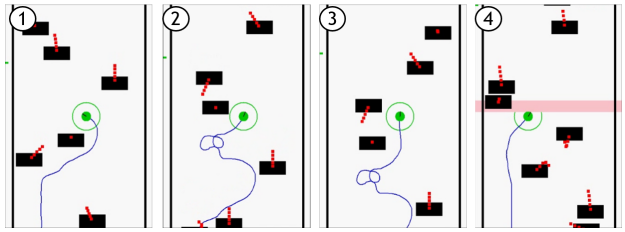



図 4: 最良個体の動作例 (提案手法 2)

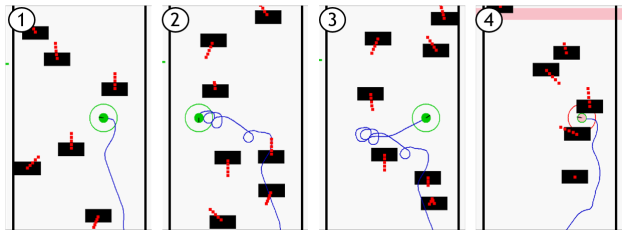



図 5: 最良個体の動作例 (RFCN)

未知環境での動作例を示す．左上の数字は時系列の順番を表す．これらの結果から，ペナルティに応じて構造の変更を行うことで，RFCN と比べて未知環境で長い間障害物を避けることができた．未知環境で事前に 5 回構造の変更を行った提案手法 2 は，事前に構造の変更を行っていない提案手法 1 と比べ最良個体では性能が良かった．しかし，10 試行の平均では変化がなかった．このことから動作中の構造変更は障害物回避の精度を向上させたが，新しい環境で事前に行う構造変更は安定性がなかった．今回の実験では，2 種類のペナルティを用意したが，主に障害物に接近したことによるペナルティが与えられていた．障害物に接近するという短期的な行動に与えられたペナルティに対して構造変更を行っていたことが原因で，環境に合わせてネットワーク構造を変えることが難しかったと考えられる．

## 4 まとめ

本稿では動作中に構造の変更を行うネットワークを提案し，障害物回避問題に適用した．その結果，未知環境で障害物回避の性能が向上した．また，事前に環境に合わせた構造の変更を行うことで最良個体の性能は向上したが，安定性はなかった．今後は長期的なエージェントの動きを考慮した構造変更を行うため，ペナルティの種類と利用するタイミングの再検討を行いたい．

## 参考文献

[1] 白川真一，長尾智晴，”RFCN による連続値空間上での自律エージェントの行動制御”, IEEJ Trans.EIS, Vol.127, No.5, pp.762-769, 2007.