2U-3

適応的に変化する進化型ニューラルネットワークによる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，知的な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一つとしてニューラルネットワークが挙げられる．筆者らの研究グループでは，ニューラルネットワークの構造や結合荷重を進化的に学習する，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．しかし一度学習で得られた構造を変更する仕組みはない．そのため環境変化などによって現状の構造で対応できなくなった場合，再学習を行う必要がある．本稿では，RFCN に調整用のフィードフォワードニューラルネットワーク (FFNN) を加えた．実験では障害物回避問題に適用し性能の検証を行った．

2　提案手法

提案手法では，RFCN にエージェントの行動中も学習を行う調整用の FFNN を加えた．これによって既存の行動を保存しつつ，新たな環境に適応することを期待している．図 1 に提案手法の概略図を示す．FFNN は 2 種類の役割の出力ユニットをもつ．1 つ目は調整するかどうかの判断を行う判断用ユニットである．2 つ目は実際に RFCN の調整を行うための調整用ユニットである．判断用ユニットが発火した場合，調整用ユニットが RFCN の出力ユニットへ接続され，出力値の調整を行う．FFNN はエージェントの行動中に，行動に応じて与えられるペナルティをもとに学習を行う．ペナルティが与えられた場合，その直前の行動は良くないと仮定し学習を行う．判断用ユニットは発火するように学習を行う．調整用ユニットは RFCN の出力が発火していた場合抑制するように，発火していない場合発火するように学習を行う．FFNN が調整を行い RFCN の出力値を変えることで，現在とは異なる行動をすることを期待している．また RFCN の出力値を調整したが一定ステップの間ペナルティが与えられなかった場合は，調整を行わないように学習を行う．調整するタイミングを限ることで，もとの行動を必要以上に変更しないことを期待している．FFNN の学習には backpropagation を用いた．学習率 α_t は毎ステップ更新を行う．式 (1), (2) に更新に用いた式を示す．

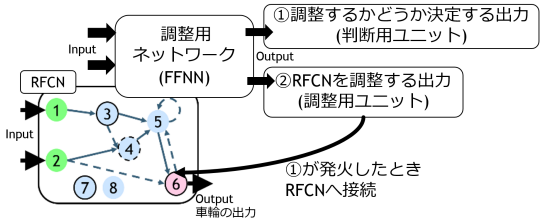

図 1: 提案手法

$$\alpha_t = P_t \alpha_0 \quad (1)$$

$$P_t = (1 - A)P_{t-1} + B(1 - P_{t-1})penalty_t \quad (2)$$

ここで，$penalty_t$ は t ステップ目にエージェントに与えられたペナルティを表す．ペナルティが与えられた場合は 1, 与えられなかった場合は 0 とする．A, B は定数である．学習率 α_t はペナルティが長期間与えられなかった場合小さくなる．これによって，ペナルティが長期間与えられない場合は現状の構造に問題はないとして，FFNN の変更が抑制される．

3　実験設定

実験では，障害物回避問題を扱った．図 2 に実験環境を示す．この実験では，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タスク

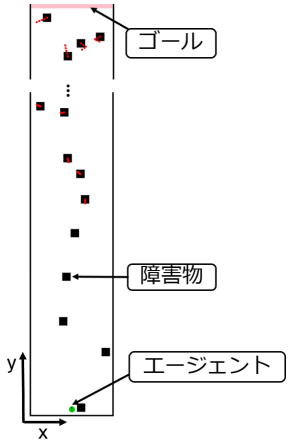

図 2: 実験環境

Action Control of Autonomous Agents using Evolutionary Neural Network Adapted to Environmental Change
†Kanako Hiroi　†Tomoharu Nagao
†Graduate School of Environment and Information Sciences, Yokohama National University

達成とし，障害物や壁にぶつかった場合はそこでタスク失敗とした．また，エージェントの初期位置に近い障害物は動かず，ゴールに近くなるほど障害物の移動速度および配置数が大きくなる．エージェントは入力情報として，9 方向の距離センサの情報と行動に対するペナルティが与えられる．ペナルティは，障害物がエージェントの一定範囲内に存在する場合に与えられる．実験では RFCN に FFNN を加えたネットワークと，通常の RFCN の比較を行った．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 (3)，(4) に示す．

$$Fitness = \frac{\sum^{\text{マップ数}} Fitness_{map}}{(\text{マップ数})} \quad (3)$$

$$Fitness_{map} = (y \text{ 軸方向の移動距離}) + (\text{ゴール報酬}) \quad (4)$$

4 実験結果

表 1 に，学習環境での適応度と未知環境 1000 マップに適用した場合の平均適応度を示す．括弧内の数値は，未知環境 1000 マップに適用した際にゴールした回数である．学習環境では RFCN のほうが有効であるが，未知環境に適用する場合は調整用の FFNN を加えたほうが適応度が上がった．また，図 3，4 に最良個体の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．どちらの手法でも障害物に近づいたとき，回転することで衝突を避ける動作が見られた．今回は適応度として距離しか考慮していないため，止まるよりも動いているほうが障害物の動きに対応し易いためだと考えられる．最良個体では FFNN を追加した場合，通常の RFCN に比べ蛇行動作が多く見られた．蛇行動作は障害物の配置が未知の環境で，ある障害物を回避しているときに別方向から来た異なる障害物の回避などに用いられていると考えられる．他にも，FFNN の調整を障害物が多い場所での車輪の減速などに用いている個体も見られた．このように FFNN の調整は，障害物に近いときの回避行動の補助として利用されていた．しかし調整ユニットの値に大きな変化は無く，調整するかどうかで行動をわけていた．場面に応じた回避行動の獲得のために，調整用ユニットの学習方法についての検討が必要である．

	提案手法		RFCN	
	学習	未知	学習	未知
平均	5132	3316 (182 回)	5270	2683 (74 回)
最良個体	5992	4980 (547 回)	5992	3914 (215 回)

表 1: 適応度 (10 試行)

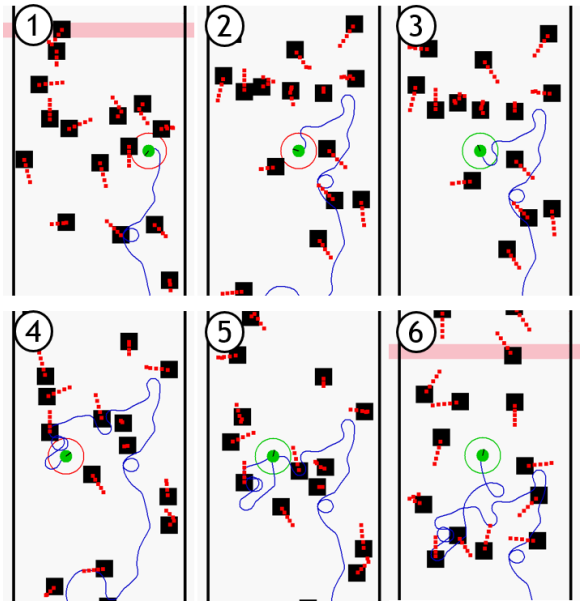

図 3: 最良個体の動作例 (提案手法)

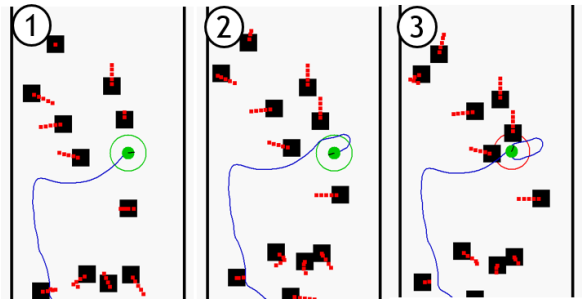

図 4: 最良個体の動作例 (RFCN)

5 まとめ

本稿では調整用 FFNN を加えた RFCN を障害物回避問題に適用した．その結果，学習時における適応度は下がるが，未知環境に適用した場合の適応度は上昇することを確認した．今後は学習時の適応度上昇の妨げとなる要因の解析と改善を行う．また，回転行動を抑制するために適応度関数にエネルギー効率の項を入れるなど，検討を行いたい．また，現状ではペナルティが与えられたときに，直前の出力値を逆方向に調整するようにしている．この調整方法については，場面に応じて異なる回避行動を獲得するために，再検討が必要である．

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.