強化学習を通したクリーチャによる協調行動の自動獲得

Automatic Acquisition of Cooperative Behavior by Creatures through Reinforcement Learning

高田 亮介 [1*] 竹内 勇剛 [1]
Ryosuke TAKATA[1] Yugo TAKEUCHI[1]

[1] 静岡大学情報学部
[1] Faculty of Informatics, Shizuoka University

Abstract: 人と機械の協調を実現するためには，機械が人の意図を推定して行動する機構をボトムアップに構築する必要がある．本研究では，意図推定を通した人と機械との協調的なインタラクションを実現するために，強化学習を用いてボトムアップに行動決定過程プロセスを構築した．実験として，箱を持ち上げる課題を 2 体のクリーチャに強化学習させた．実験の結果，ニューラルネットワークから行動原理の説明が可能なクリーチャが得られ，クリーチャ間にリーダーとフォロワーの関係が見られた．今後は，クリーチャと人との協調に向けた検討を行っていく．
キーワード 強化学習，協調，リーダー・フォロワー，ニューラルネットワーク

1 はじめに

引っ越し作業では，重たい荷物を二人，または三人で運ぶことが多い.「せーの」などと声をかけ，タイミングを合わせて持ち上げる．このような，複数人で互いに調整しながら同じ作業を行うとき，人はごく自然に，かつ円滑に協調を行う．このような重労働はできるだけ機械に任せたい，というのが近年の社会動向であるが，機械が協調を行うことは未だ困難である．1 体の機械で持ち上げることも考えられるが，重たい荷物を持ち上げられるように，機械自体を大きく丈夫に設計する必要がある．複数体で持ち上げることで，単純な力の足し合わせよりも全体として大きな力を発揮できることがあり，それこそが協調する意義である．

機械同士が協調する取り組みは,「協調荷押し課題」での強化学習による分担作業の自動獲得 [1] や,「ハンター課題」での意図推定を用いた協調行動の分析 [2][3] など，様々な課題を題材に行われている．これらの先行研究の問題点として，人と協調可能であるか，ということが挙げられる．人と協調するためには，個体の意図を推測し，行動を柔軟に調整する必要がある.「協調荷押し課題」のように群知能的なアプローチでは，個体の意図を推測する必要のある課題を達成するのは難しい．また「ハンター課題」のように，設計者が予め行動原則をトップダウンに決めてしまう方法では，柔軟な協調行動を獲得することは難しい．人と協調可能かどうか明らかにしたい場合，意図推定が可能かどうか，という点が重要であると考えられる．さらに，人が行動ルールを設計しないボトムアップなアプローチで行動を決定させる必要がある．本研究では，意図推定と強化学習を用いて協調行動を自動獲得する実験を行う．実験環境として，2 体のクリーチャが 1 つの重たい箱を持ち上げる「箱持ち上げ課題」を用意した．学習手法には，ニューロエボリューションの代表的な手法である NEAT を用いた．これにより，他の強化学習手法より短時間で最適な行動とニューラルネットワークを得られるため，協調行動における行動決定プロセスを分析できる．本研究の目的は，機械に対して協調行動をボトムアップに獲得させ，その行動原理を分析することである．将来的には，人の意図を汲み取って行動する機械の実現させ，人と機械が円滑に協調するための手法を確立したい．

*連絡先： 静岡大学情報学部情報科学科
〒 432-8011 静岡県浜松市中区城北 3-5-1
E-mail: cs16503@s.inf.shizuoka.ac.jp

2 背景

2.1 強化学習による協調行動の獲得

大倉ら (2011) は，自由に動き回ることのできる 2 次元環境で，10 体のロボットによる協調荷押し課題を強化学習させている [1]．大きさの異なる 3 つの荷物は運ぶために必要なエージェント数が決まっており，この分担が協調であると定義して実験を行っている．この研究では，強化学習手法に NEAT が用いられており，

NEAT によって協調的振る舞いが獲得されたことが報告された．ただし，本研究とは協調の捉え方のアプローチが異なる．大倉らの先行研究が群知能的アプローチによって全体の最適化を図るのに対し，本研究では，個体の振る舞いや行動原理であるニューラルネットワークに注目した．ニューラルネットワークを分析することで，より個体レベルの動作に焦点をあてた分析を行う．

椿本ら (2015) は，Q 学習を用いて 2 次元グリッド空間上の荷物運び課題を強化学習させている [4]．この研究により，意図推定法を導入した方が協調行動が成功しやすいことが示唆された．

佐藤ら (2007) は，侵入ゲームと呼ばれる 4 マスの簡単なゲームにおいて，コミュニケーション手段としての「発光」を導入した協調課題を強化学習させている [5]．この研究により，発光行動は互いに共有された信号として，意味を創発させ得る可能性が示唆されている．本研究においても，他者が知覚可能な発光を行えるようにする．発光によって他者に意図を伝達し，相手の推定した意図の正確性が向上して円滑に協調を行えるようになるのではないかと考えられる．

2.2 意図推定レベル

人は，他者の意図を推測しながら行動している．ここで，意図推定には心の理論に基づく再帰のレベルが存在することが知られている [2][6]．表 1 に，意図推定レベルの一部を示す．なお，表 1 ではレベル 2 までしか記述していないが，意図推定レベルは無限に再帰し得る．

表 1: 行動主体が推定する意図推定のレベル

レベル	定義
0	他者の意図を推定せず，自己の意図にのみしたがって自己の行動を決定する行動主体
1	他者をレベル 0 と想定してその意図を推定し，予測される他者の行動に対応して自己の行動を決定する行動主体
2	他者をレベル 1 と想定して他者が推定する自己の意図を推定し，その推定から予測される他者の行動に対応して自己の行動を決定する行動主体

2 体の行動主体が相互に相手を予測する場合，同じ行動決定過程，つまり意図推定レベルを有していると円滑な協調が行えないことが知られている [3]．この問題は相互予測問題と呼ばれる．相互予測問題は，意図推定のレベル差が 1 のときに解決されることが明らかになっている．

3 実験方法

3.1 協調課題

椿本ら (2015) は，協調を「他者と同じ目標を持って行動すること」と定義している [4]．今回は，クリーチャ 1 体では持ち上げることができない重たい箱を，2 体のクリーチャが持ち上げる課題を用いて実験した．実験は図 1 に示すような 2 次元仮想環境上で行った．環境には物理計算ライブラリを使用しているため，物理法則に従って挙動する．図 1 において，緑の円形がクリーチャ（左がクリーチャA，右がクリーチャB），赤い長方形が持ち上げる対象である箱，青の水平線が課題達成ラインである．なお，クリーチャの初期位置は箱の下でランダムに決めており，エピソード毎に初期位置が異なるようにした．また，1 エピソードは 500 ステップ以内で構成され，箱の角度が $\pm 0.5[rad]$ 以上傾いた時点でそのエピソードは強制終了されるように設計した．

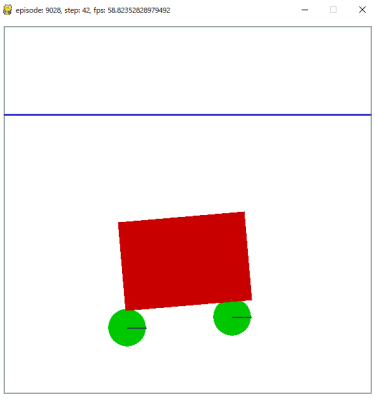


図 1: 2 体のクリーチャによる箱の持ち上げ課題

3.2 実験条件

ニューロエボリューションの代表的な手法である NEAT[7] を用いて，表 2 に示す 5 条件の実験を行った．表 2 におけるそれぞれの条件要素については，以下で説明する．

表 2: 実験条件

条件	発光	NN	状態空間
C-1	×	FNN	同型
C-2	○	FNN	同型
C-3	○	RNN	同型
C-4	○	RNN	意図推定レベル
C-5	○	RNN	片方発光のみ

3.2.1　発光

条件 C-2 以降では，クリーチャが「発光」可能な状態で実験を行った．ここで，発光とは，クリーチャが認識することのできる 2 値のシグナルを発信する手段として設計した．設計者は発光自体に意味を与えず，クリーチャが自ずと意図を伝達する手段として発光を用いることを期待して導入した．なお，移動（右移動，左移動，ジャンプ）しながら発光を行うことができるようにした．今回は，発光とニューロンの出力値の関係を，以下の式 (1) のように定義した．式 (1) において，p は出力ニューロンの値で，$light$ が 0 の場合は発光しない状態，1 の場合は発光する状態となる．

$$light = \begin{cases} 0 & (p \leq 0) \\ 1 & (p > 0) \end{cases} \tag{1}$$

3.2.2　ニューラルネットワーク

条件 C-2 以前ではニューラルネットワークに FNN を用い，条件 C-3 以降では RNN を用いて実験を行った．FNN は単一方向にのみ情報を伝播するネットワーク構造 [8] で，RNN は再帰的な伝播を含むネットワーク構造 [9] である．条件 C-2 と C-3 の比較により，FNN と RNN のどちらが今回の協調課題に適しているか検証した．RNN は前ステップの行動の時系列，つまり文脈を考慮することができるため，今回扱っているような連続時間の中で行われる協調課題においては，FNN より RNN を用いた方が課題成功率が高くなると予想した．

3.2.3　状態空間

2 体のクリーチャにおける状態空間の組み合わせを 3 種類用意した．「同型」は，状態空間によってクリーチャに役割を与えないように，2 体のクリーチャに同じ状態空間を与えて実験した条件である．「意図推定レベル」は，意図推定レベルを設計者が導入した条件である．具体的には，意図推定レベル 0 のクリーチャは相手クリーチャの位置を観測せず，箱の位置だけを観測して行動を決定する．そして意図推定レベル 1 のクリーチャは，意図推定レベル 0 の相手クリーチャの位置を観測して行動を決定する．「片方発光のみ」は，クリーチャA には全ての情報を観測できるように設計し，クリーチャB には発光の情報のみを観測できるように設計した条件である．クリーチャB は発光以外の環境情報（箱や相手の位置情報）を把握することができないため，協調課題を成功させるために，クリーチャA はクリーチャB に対して発光を用いることで，環境に関する情報を伝達するようになるのではないかと予想した．

4　学習結果

NEAT を用いた学習結果の一覧を表 3 に示す．以下，各条件ごとに詳細に結果を説明する．

表 3: 学習結果一覧

条件	発光	NN	状態空間	結果
C-1	×	FNN	同型	◯
C-2	○	FNN	同型	◯
C-3	○	RNN	同型	◎
C-4	○	RNN	意図推定レベル	◯
C-5	○	RNN	片方発光のみ	△

4.1　条件 C-1: 発光無し

条件 C-1 においては，課題成功率はおよそ 50[%] 程度であった．このときの適合度の推移を図 2 に示す．また，この学習によって得られた各クリーチャのニューラルネットワークを図 3 に示す．

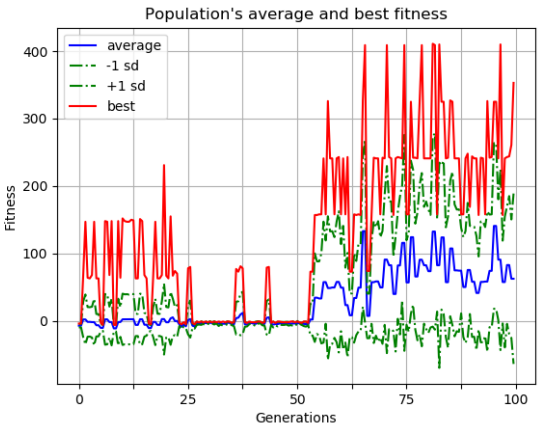
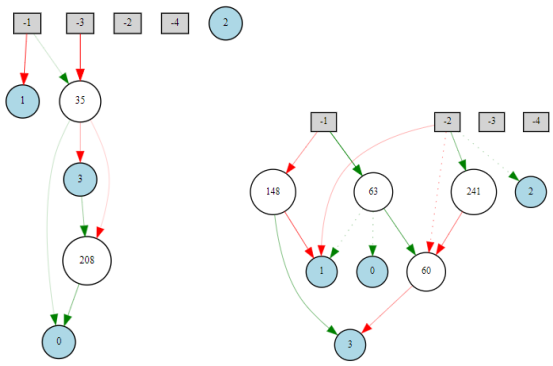


図 2: 条件 C-1 における適合度の推移



(a) クリーチャA　　(b) クリーチャB

図 3: 条件 C-1 におけるニューラルネットワーク

4.2 条件 C-2: 発光の導入

条件 C-2 における適合度の推移を図 4 に示す．また，この学習によって得られた各クリーチャのニューラルネットワークを図 5 に示す．図 4 のグラフは右肩上がりで，100 世代まで適合度が上昇していることがわかる．学習を続けることで，さらに適合度が上がることも予想される結果だった．クリーチャの振る舞いに注目すると，2 体とも課題開始から終了まで発光を行うことは無く，開始直後からすぐに箱を持ち上げ始めて，位置を微調整しながら課題に成功する場面が多く見られた．

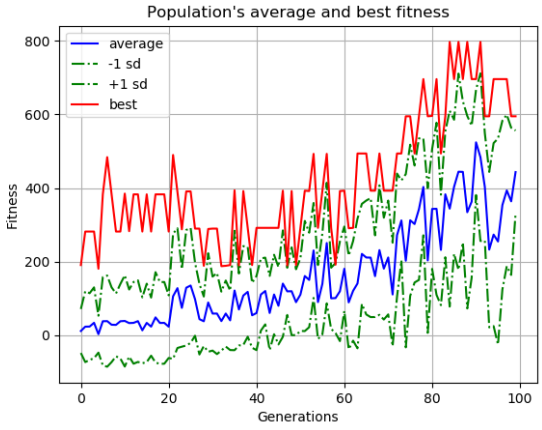
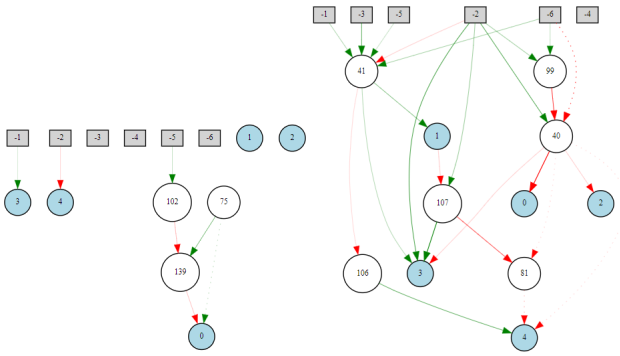


図 4: 条件 C-2 における適合度の推移



(a) クリーチャA　(b) クリーチャB

図 5: 条件 C-2 におけるニューラルネットワーク

4.3 条件 C-3: RNN の導入

条件 C-3 における適合度の推移を図 6 に示す．図 6 より，他の条件と比べて飛躍的に適合度が高いことが分かる．クリーチャの振る舞いを確認すると，クリーチャA，B 共に相手との位置関係を微調整しながら箱をバランスよく持ち上げる様子が見られた．このときの様子を図 7 に示す．目視による観測ではほとんど失敗せず，どんな初期位置でも課題を成功していた．さらに注目すべきは，課題開始直後は互いにジャンプを行わず地面を沿うように移動しており，クリーチャB が発光し始めたら互いに箱を持ち上げ始めて課題成功していたことである．このとき，クリーチャA は課題開始から終了まで常に発光し続けており，クリーチャB は課題開始直後は発光せず通常状態で，クリーチャ間の距離がある程度近づいたら発光していた．また，学習の結果得られたクリーチャのニューラルネットワークを図 8 に示す．図 8 の各ニューラルネットワークには，RNN の特徴であるニューロンの再帰構造が含まれていることが確認できる．

条件 C-3 の結果は全条件の中で最も高い適合度であったため，さらにクリーチャB をランダムに行動させるように設計し，学習によって得られたクリーチャA と課題を行わせた．その結果，ランダムに細かく移動するクリーチャB に対して非常に敏感に反応し，バランスのとれた位置関係を保つように持ち上げていた．このように，ランダムに行動するクリーチャに対しても協調行動を実現することができた．

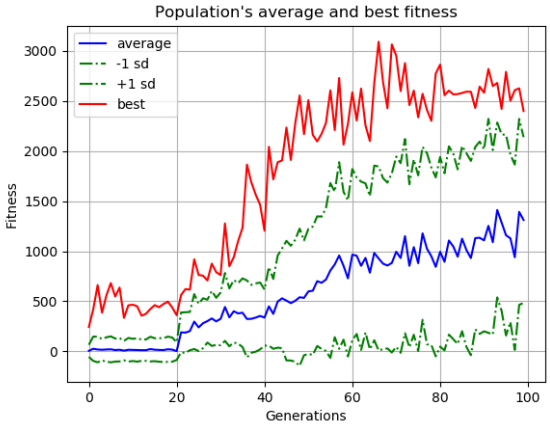


図 6: 条件 C-3 における適合度の推移

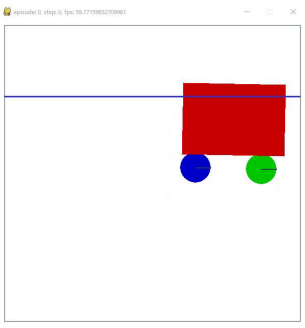
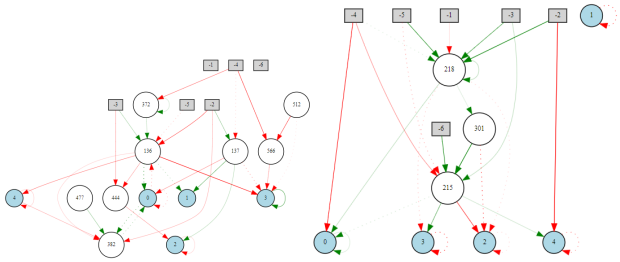


図 7: 条件 C-3 における協調課題成功の様子



(a) クリーチャA　　(b) クリーチャB

図 8: 条件 C-3 におけるニューラルネットワーク

4.4　条件 C-4: 意図推定レベル

条件 C-4 における適合度の推移を図 9 に示す．また，学習の結果得られたクリーチャのニューラルネットワークを図 10 に示す．クリーチャA は意図推定レベル 1，クリーチャB は意図推定レベル 0 として状態変数を設定した．図 10(a)（意図推定レベル 1）の方が，図 10(b)（意図推定レベル 0）よりも複雑な構造になっていることが分かる．

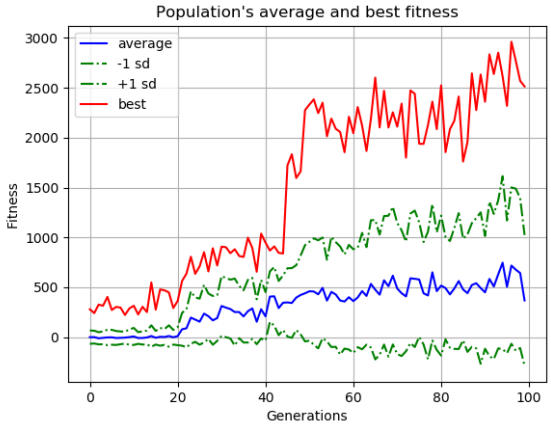
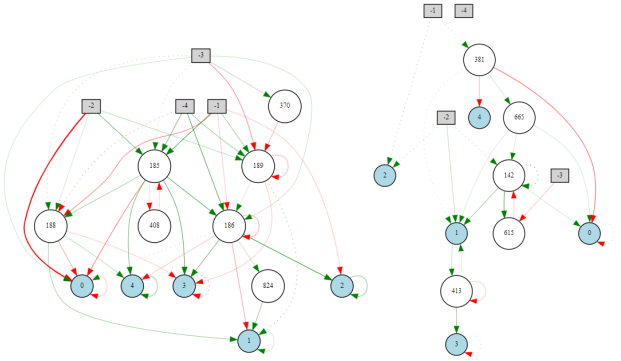


図 9: 条件 C-4 における適合度の推移



(a) クリーチャA（レベル 1）　　(b) クリーチャB（レベル 0）

図 10: 条件 C-4 におけるニューラルネットワーク

4.5　条件 C-5: 情報不足設計

条件 NEAT-8 における適合度の推移を図 11 に示す．また，学習の結果得られたクリーチャのニューラルネットワークを図 12 に示す．まず，図 12(a) のクリーチャA のニューラルネットワークに注目すると，入力 1（相手の相対距離），入力 2（相手の相対角度），入力 3（箱の相対距離）の環境情報を集約している中間ニューロン 207 番から，出力 4（発光）に信号が伝播する構造になっていることが分かる．次に，図 12(b) のクリーチャB のニューラルネットワークに注目すると，入力 1（自分の発光状態），入力 2（相手の発光状態）は他のニューロンと結合してなく，信号が伝播しない構造であることが分かる．クリーチャB は，中間ニューロン 846 番からのバイアス値を正の結合荷重を通して伝播し，常に出力 3（ジャンプ）が発火するような行動になっている．

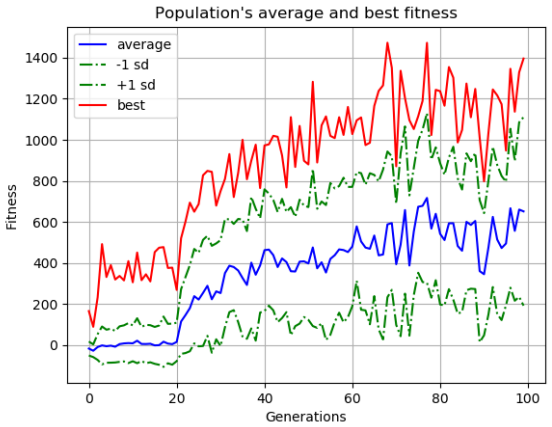
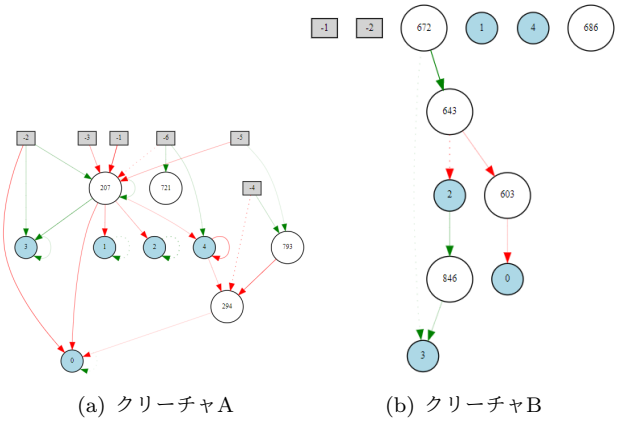


図 11: 条件 C-5 における適合度の推移



(a) クリーチャA　　(b) クリーチャB

図 12: 条件 C-5 におけるニューラルネットワーク

5 考察

5.1 意図推定レベルの出現（C-1，C-2）

条件 C-1，C-2 の結果から，意図推定レベルが状態変数として自動的に獲得されたことが考えられる．条件 C-1 の場合，図 3(a)（クリーチャA のニューラルネットワーク）からは，入力 1（相手の距離）及び入力 3（箱の距離）のみを伝播して行動を決定していることが分かる．これは，相手と箱の大まかな情報から行動するリーダーであると考えられる．また，図 3(b)（クリーチャB のニューラルネットワーク）からは，入力 1（相手の距離）及び入力 2（相手の角度）のみを伝播して行動を決定していることが分かる．これは，相手の情報のみをもとに行動を決定するフォロワーであると考えられる．条件 C-2 の場合も C-1 と同様に（発光状態に関する入力以外に注目すると），図 5(a) より，クリーチャA は入力 1（相手の距離）及び入力 2（相手の角度）のみを伝播して行動を決定しているフォロワー，図 5(b) より，クリーチャB は入力 1（相手の距離），入力 2（相手の角度），入力 3（箱の距離）を伝播して行動を決定しているリーダーであると考えられる．

ここで注目したいことは，条件 C-1 では意図推定レベルを設計者が導入したわけではなく，クリーチャがボトムアップに構築したことである．つまり，入力は設計者によって制限されていないにも関わらず，進化的な学習の結果，自動で得られたニューラルネットワークが意図推定レベルを表す制限された状態変数を伝播する構造になっていたのである．以上より，意図推定レベルの構造を自動獲得できる可能性が示唆される．

5.2 発光の有無（C-1，C-2）

条件 C-1 と C-2 の比較より，発光可能であるかどうかによって行動の変化は見られなかった．この結果は，発光を行わなくても協調課題が成功する，ということを表していると考えられる．しかし，図 2 と図 4 の比較から，発光可能なクリーチャの方が最終的な適合度が高いことがわかる．つまり，発光を実際に行うかどうかより，発光を行うことのできる状態であることが，適合度の上昇に繋がったのではないかと考えられる．具体的には，図 5(a)（クリーチャA のニューラルネットワーク）に注目すると，入力 2（相手の角度）を観測して出力 4（発光）を決定している構造であり，さらに，入力 5（自身の発光状態）から出力 0（静止）を決定している．この構造により，再帰性を含んでいると解釈することができる．あるステップで入力 2（相手の角度）が閾値を超えると発光を行い，その結果，次のステップで入力 5（自身の発光状態）が変化し，静止する．すなわち，発光は他者に対するシグナルであるだけでなく，自身に対するシグナルでもあると考えられる．

5.3 発光への意味の創発（C-3）

図 7 より，条件 C-3 では，発光に意味を創発したと考えられる．結果より，課題開始直後は互いに距離を調整するように移動し，一定の距離まで近づいてバランス可能な状態になったらクリーチャB が発光し，そのタイミングで同時に箱を持ち上げ始める．この一連の動作から，発光は「持ち上げ始める合図」としての意味を持っていると解釈できる．図 8(b) から，合図を出すクリーチャB は，環境情報を集約した中間ニューロン 215 番と，入力 2（相手の相対角度）をもとに出力 4（発光状態）の値を決定していることが分かる．さらに，出力 3（ジャンプ）への伝播も出力 4（発光状態）と似た構造になっているため，発光すると同時にジャンプを行うような仕組みになっていると考えられる．しかし，ここで図 8(a) に注目すると，クリーチャA は入力 6（相手の発光状態），つまりクリーチャB の発光状態は行動決定プロセスに組み込まれていないことが分かる．これにより，クリーチャB は合図を出しているが，クリーチャA はその合図を内部的には受け取っておらず，相手との位置関係のみで相手に合わせるように行動していると考えられる．その結果，発光が両者の間にあたかも「持ち上げ始める合図」としての意味が創発されたかのように考えられる．

5.4 ニューラルネットワーク（C-2，C-3）

NEAT に用いるニューラルネットワークを FNN から RNN に変更することで，明らかに適合度の水準が上昇したことが，図 4 と図 6 の結果から分かる．RNN は時系列を扱うことが可能であり，1 ステップ前の行動から現在の行動を決定するといった文脈の記憶機能がある．今回の箱持ち上げ課題のように，繰り返しパターンによって達成される課題においては，RNN は非常に有効であると考えられる．

5.5 状態空間による関係形成（C-3，C-4，C-5）

2 体のクリーチャに異なる意図推定レベルを表す状態空間を設定した条件 C-4 の結果から，図 10(a)（意図推定レベル 1 であるクリーチャA のニューラルネットワーク）の方が，図 10(b)（意図推定レベル 0 であるクリーチャB のニューラルネットワーク）よりも中間ニューロンが多く複雑であることがわかる．これより，意図推定レベル 1 は相手の動作を考慮して行動するた

め，多くの中間ニューロンを経由することでその複雑なプロセスを構築しているのではないかと考えられる．

また，クリーチャBの入力を発光情報のみに設計した条件C-5の結果では，図12(b)のクリーチャBのニューラルネットワークは比較的シンプルな構造になっていることが分かる．ここで注目すべきは入力値がどのニューロンにも伝播しない構造になっていることである．共進化的な学習の結果，発光情報しか与えられないクリーチャBは入力値に頼った行動を妥協し，常にジャンプし続けることで課題成功するようになったと考えられる．この背景には，クリーチャAがクリーチャBに合わせるように行動できる図12(a)のニューラルネットワークが必須であり，クリーチャBがひたすらジャンプし続ける予測可能な環境の一部となったからこそ，クリーチャAが協調的に振る舞うことが可能となったと考えられる．

5.6 人との協調可能性と解決すべき問題

条件C-3で学習したクリーチャがランダムに行動するクリーチャとの協調に成功した結果から，人との協調も十分に可能であると考えられる．ただし，協調を行える範囲は本研究で取り扱った「箱持ち上げ課題」内で閉じており，汎用的な協調が可能になったとは言えない．課題の外側でやり取りされる情報（例えば相手の身体的な疲労状態や相手のこれからの予定などの事前情報）が不足していることで，適切で柔軟な行動判断が行えないことがあり，実際に現実世界で人と協調する際には，この問題を解決する必要がある．

さらに解決すべき点として，視点レベルの共有が挙げられる．人はメタな視点を持っているため，課題をマクロに捉えることができる．将来的に人と機械との協調を考えると，機械も人のようにマクロに捉える視点が必要であると考えられる．意図推定は，相手の意図を自身行動決定プロセスの一部とするが，自身がマクロな視点を持っていない場合，相手のマクロな視点に対応して行動することは困難であると考えられる．よって，人と機械の協調を実現する上で，人と同じ視点のレベルで状況を捉える仕組みは必要不可欠であると考えられる．

6 おわりに

本稿では，クリーチャ同士の協調行動の自動獲得を目指して実験を行った．具体的には，2次元空間上での箱の持ち上げ課題についてNEATを用いて強化学習させた．学習の結果，柔軟に振る舞う協調的な行動をボトムアップに獲得できた．クリーチャの振る舞いに注目すると，意図推定レベルから形成されるリーダーとフォロワーの関係が出現したり，発光に意味を創発したかのように行動する結果が得られた．さらに，学習によって得られたニューラルネットワークから，協調行動決定プロセスがある程度説明可能であることを示した．このように，協調行動原理が明確化されることで，人に対しても協調系を築くことが可能であると考えられる．

今後は，人による操作を可能にしたいと考えている．学習の結果得られたクリーチャの行動は，非常に柔軟な適応力を持ち合わせていると考えられる．特に条件C-3でのクリーチャAは，ランダムに行動するクリーチャとも協調行動が成功できた結果から，クリーチャAと人との協調も成功するのではないかと考えられる．そこで，今後は人がクリーチャBを操作できるようにして，人とクリーチャの協調を実現したい．そして将来的には人とクリーチャが自然に協調できる手法を確立させたい．

参考文献

[1] 大倉和博, 保田俊行, 松村嘉之: 構造進化型人工神経回路網による Swarm Robotics のための適応的協調行動の生成, 日本機械学会論文集（*C* 編）, Vol.77, No.775, pp.399-412 (2011).

[2] Yugo Nagata, Satoru Ishikawa, Takashi Omori, Koji Morikawa: Computational model of cooperative behavior: Adaptive regulation of goals and behavior, *Proceedings of Second European Cognitive Science Conference*, pp.202-207 (2007).

[3] 長田悠吾, 石川悟, 大森隆司, 森川幸治: 意図推定に基づく行動決定戦略の動的選択による協調行動の計算モデル化, *Cognitive Studies*, Vol.17, No.2, pp.270-286 (2010).

[4] 椿本樹矢, 小林邦和: 意図推定法を用いたマルチエージェント強化学習システムにおける協調行動の獲得, 電気学会論文誌 *C(*電子・情報・システム部門誌*)*, Vol.135, No.1, pp.117-122 (2015).

[5] 佐藤尚, 内部英治, 銅谷賢治: 強化学習エージェントによる協調行動とコミュニケーションの創発, *TOM*, Vol.48, No.19, pp.55-67 (2007).

[6] 高野雅典, 加藤正浩, 有田隆也: 心の理論における再帰のレベルの進化に関する構成論的手法に基づく検討, *Cognitive Studies*, Vol.12, No.3, pp.221-233 (2005).

[7] Kenneth O. Stanley, Risto Miikkulainen: Evolving neural network through augmenting topologies, *Evolutionary Computation*, Vol.10, pp.99-127 (2002).

[8] Christopher M. Bishop: Neural Networks for Pattern Recognition, *Oxford University Press* (1995).

[9] Gintaras V. Puskorius, Lee A. Feldkamp: Neurocontrol of nonlinear dynamical systems with kalman filter trained recurrent networks, *IEEE Transactions on Neural Networks*, Vol.5, pp.279-297 (1994).