様式 C-19

科学研究費補助金研究成果報告書

平成２３年　５月２０日現在

機関番号：１７５０１
研究種目：基盤研究(B)
研究期間：2007～2010
課題番号：１９３０００７０
研究課題名（和文）　ニューラルネットを用いた強化学習で、どこまで高次機能の創発が説明できるかへの挑戦
研究課題名（英文）　A challenge towards how far the emergence of higher functions can be explained by reinforcement learning using a neural network
研究代表者
柴田　克成（SHIBATA KATSUNARI）
大分大学・工学部・准教授
研究者番号：10260522

研究成果の概要（和文）：
センサからモータまでをニューラルネットで接続し、強化学習による自律的な学習によって「高次機能創発」を目指した。一つの柱である「シンボル処理」の創発に関しては大きな成果を示すことができなかったが、報酬獲得のためには画像上の矢印の向きの判別と記憶が必要な可動カメラを用いたタスクで、画像から向きの抽出と記憶を実現できたことを始め、単なる報酬や罰からの学習で「抽象化」「記憶」「予測」「探索」などの高次と呼べる機能の創発を示した。

研究成果の概要（英文）：
"Emergence of higher functions" has been aimed based on autonomous learning by reinforcement learning using a neural network that connects from sensors to motors. Although no significant fruits could be shown as for the emergence of symbol processing, but the emergence of "abstraction", "memory", "prediction", or "exploration" through the learning from only rewards and punishments has been shown. For example, in the learning of a task using a movable camera in which the identification and memorization of arrow direction are necessary to get a reward, extraction of arrow direction from images and memorization of it were realized.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2007 年度	2,600,000	780,000	3,380,000
2008 年度	1,100,000	330,000	1,430,000
2009 年度	1,200,000	360,000	1,560,000
2010 年度	1,100,000	330,000	1,430,000
年度			
総　計	6,000,000	1,800,000	7,800,000

研究分野：総合領域
科研費の分科・細目：情報学、知覚情報処理・知能ロボティクス
キーワード：知能ロボット、強化学習、ニューラルネット、高次機能

１．研究開始当初の背景

従来、ロボットを高機能化させるために、ロボットのプロセスを、認識、プランニング、制御といった機能モジュールに分割し、それぞれの機能が高機能になるように設計者がプログラミングして、全体として高機能にするというアプローチがとられてきた。これに対し、筆者らは、ロボットが自ら学習して賢くなる枠組みとして、「センサからモータまでを分断せずに一つのニューラルネットで

構成し、強化学習で自律的に学習することで、内部に必要に応じた機能が創発する」というアプローチを提唱し、ロボットが、報酬や罰に基づいて、カメラからの視覚情報を入力として適切に箱押し行動することを試行錯誤に基づく学習で獲得きることなどを示して来た。

人間における「高次機能」の解明とともに、ロボットの世界で「高次機能」を実現することは長年の課題であった。しかし、基本的には前述のモジュール分割の考え方の延長線上で捉えられて来ており、結局タスクごとに設計者としての人間が高機能なプロセスを設計することに留まり、ロボットにおける運動機能の向上と比較して、「高次機能」の方はあまり進展がない状況が続いていた。

２．研究の目的

「高次機能」は、センサに近い「認識」やモータに近い「制御」とは異なり、そもそも何を入力として何を出力とすべきかを予め決めることも難しいものである。したがって、一見遠回りに見えても、筆者らが提案して来た、センサからモータまでをニューラルネットで接続し、システム全体を強化学習で自律的に学習させるという枠組みが有効であるとの考えに基づき、学習による「高次機能の創発」を目標とした。さらに、実ロボットを用いたデモを行なうことも目標とした。

そして、「視覚センサを有する実移動ロボットを、いくつかの部屋からなる迷路の中に置いて、多数の視覚センサ信号を直接ニューラルネットへ入力して報酬を得る行動を強化学習によって獲得させる」という具体的なタスクを設定し、以下の３つのサブテーマに分けて研究を進めることを目指した。

(1) 離散状態遷移の学習とシンボル・論理的思考

今どの部屋にいるかという離散的な状態表現を自ら獲得するとともに、三段論法のような論理的思考への可能性を探る。

(2) 空間情報の抽象化と予測・概念形成

正しい道順を示す矢印の画像を見て行動を学習することで、矢印の向きが重要であることを発見し、矢印の向き、つまり、抽象化された情報を抽出して記憶し、正しい行動に結びつけることが学習できるかを検証する。さらに、学習を通して獲得された入力情報の抽象化によって、学習結果が他の矢印の場合にどれくらい汎化されるかを調べていく。

(3) 決定論的知的探索と時間的抽象化・好奇心

「学習による探索」がどこまでできるかを見極めるとともに、「未知能状態を知る」ことの価値や「好奇心」に基づく探索行動を検討する。

３．研究の方法

(1) 離散状態遷移の学習とシンボル・論理的思考

リカレントニューラルネットの構造と乗算ニューロンの導入による離散状態間遷移の学習の性能への影響を、時間軸に展開したEXOR 問題と、数を数えるカウンタータスクで調べた。

また、当初購入したロボットを用いて、複数の部屋からなる環境で学習させる予定であったが、購入したロボットが故障続きで、結局本研究期間内で想定していた環境の構築はできなかった。そこで急遽、シミュレーションによって、スイッチを踏み、その際に得られる信号によって指示されたゴールに向かうと報酬が得られるというタスクの学習を行なった。そして、スイッチを踏む前の状態から踏んだ後の状態への遷移とリカレントニューラルネットにおけるその表現の解析、さらには、入力信号による状態の分岐表現を観察した。

さらに、本研究期間後半には、当初の計画にはなかったが、「シンボル処理」の創発への糸口を探るため、論理的な思考が内在化したコミュニケーションとして捉えられるのではないかと考え、コミュニケーションの学習に着手した。具体的には、送信者は、ロボットがいるフィールドを上からカメラで捉えた画像をニューラルネットへの入力とし、２つの出力をそれぞれの周波数とする２つの音声をスピーカーから発し、受信者はその音声をマイクで拾い、FFT に掛けた信号をニューラルネットへの入力としてロボットの行動を決定し、ロボットがゴールに到達すると報酬がもらえるという設定とした。これによって、強化学習を通してコミュニケーション信号の意味付けが行なわれ、必要な情報をコミュニケーションできるようになるかどうかを確認した。

(2) 空間情報の抽象化と予測・概念形成

筆者が提案した空間情報の抽象化学習のモデルが人間でも行なわれているかどうかを調べるため、心理物理実験を行なった。

さらに、実際の可動カメラを用いて、モニタ上に表示した数パターンからランダムに選んだ一つの矢印の画像をリカレントニューラルネットに入力し、カメラ動作を決定し、矢印が見えなくなった後も矢印の方向にカメラを動かすと報酬がもらえる設定で強化学習を行なった。これによって、矢印のパターンによらずに、向きが重要であることを発見し、それを抽出できるようになるか、さらに、違った矢印パターン間での汎化能力について検証した。この実験も前述の複数の部屋よりなる環境での実験の代わりに行なった。

予測の学習については、従来は強化学習と

は別に予測を行なうモジュールを設けて、教師あり学習が行なわれて来ていたが、そもそも何を予測するのかが非常に重要な問題であることを指摘し、ここでは、可変方向、可変速度で移動し、かつ、途中で目に見えなくなる物体を捕獲するという予測を必要とするタスクを、リカレントニューラルネットを用いた強化学習で学習させることで、何を予測すべきかも含めて、内部に予測の機能が創発することを検証した。

(3) 決定論的知的探索と時間的抽象化・好奇心

学習のために探索するという従来の枠組みを 180 度ひっくり返し、学習によって知的な探索を行なうことを試みた。毎回形の異なる簡単な迷路を用意し、ゴールをランダムに配置し、エージェントから見えない状態で記憶の機能を学習できるリカレントニューラルネットを導入して、強化学習を行なった。そして、過去の探索を考慮して、迷路をくまなく探索することができるようになるかどうかを検証した。また、簡単な分岐を設けた環境で学習させることで、単に現在の状態からの静的なマッピングで探索しているのではなく、過去の状況を考慮した探索を実現できているかどうかを確認した。

最後に、前科研費の研究の続きの内容となるが、カメラを有する 4 足歩行型の実移動ロボットが、カメラの画像を元に移動し、相手のロボットとキスができたら報酬がもらえ、相手を見失ったら罰をもらう設定で、背景や照明条件を変化させながら学習を行ない、単なる報酬や罰に基づく学習で、カメラの画像から、背景や照明条件に惑わされずに相手のロボットの位置を正しく認識し、到達することができるようになるか検証した。

4．研究成果

(1) 離散状態遷移の学習とシンボル・論理的思考

まず、リカレントニューラルネットの学習において、出力からのフィードバックが学習に有効であることがわかったが、その他の学習において、一般的に出力からのフィードバックが有効であるという結果は得られなかった。また、乗算ニューロンを導入してカウンタータスクの学習をさせたところ、乗算ニューロンがない状態ではほぼ学習できなかったものが、乗算ニューロンを導入することで学習成功率が 30%程度に上昇した。しかしながら、カンタータスクが非常に簡単なタスクであることを考えると、さらなる学習能力の向上が必要である。

また、層構造を基本としたリカレントニューラルネットにおいて、層数を変化させたり、どの中間層ニューロンの値をフィードバックさせるか、また、ニューラルネット内の重み値の初期値による影響を、時間軸に展開した EXOR 問題で調べ、下の方の層でのフィードバックが有効であること、また、他のニューロンとのフィードバック結合を 0 ではなく、ある程度乱数で与えた方が学習性能が良いことがわかった。しかしながら、なぜそのようになるか、他のタスクではどうかは今後の課題として残っている。

次に、実ロボットの故障で急遽始めた、スイッチ踏みタスクのシミュレーションでは、ロボットのスタート位置、スイッチ、ゴールの位置を変化させても、２つのゴールのうち、スイッチを踏んだ際に得られる信号で指示される正しいゴールの方へ向かう行動を強化学習によって獲得することができた。その際、リカレントネットの内部を観察すると、スイッチを踏んだ状態であることを表現するニューロン群、ゴール１へ向かうことを表すニューロン群、ゴール２へ向かうことを表すニューロン群が存在することがわかった。また、行動途中にニューロンの内部状態を変化させることで、その行動を変化させることができ、スイッチへ向かう状態、ゴール１へ向かう状態、ゴール２へ向かう状態と、連続空間、連続行動の中で、必要に応じて離散的な状態表現ができることがわかった。さらに、ニューラルネットの内部状態を操作することで間違ったゴールに到達した場合は、図１

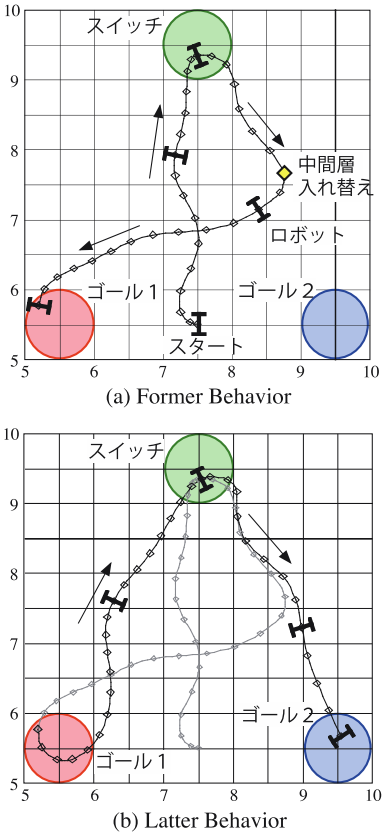

図１途中で中間層ニューロンの値を入れ替えた場合のロボットの行動

のように、再びスイッチに戻り、正しいゴールを確認し直して、正しいゴールへ向かうという知的と感じられるような行動も観察することができた。ただし、複数のスイッチを踏んでからゴールに進むというタスクの学習はできなかった。以上のことから、３つ以上の複数の状態を遷移して行くようなタスクをいかに学習させるかということが今後の課題として浮かび上がった。

また、コミュニケーションタスクでは、何をコミュニケーションさせるべきかも与えていないにも関わらず、ロボットがゴールに着いた時の報酬と壁にぶつかった時の罰だけから、ロボットが映った画像の１千個以上の信号から必要な情報を音声で送信し、受信者はその音声を聞いて、ロボットがゴールに向かうための行動を指示することができるようになった。

(2) 空間情報の抽象化と予測・概念形成

筆者が提案した空間情報の抽象化学習のモデル検証のための心理物理実験は、被験者に音声と視覚を組み合わせた新たな抽象化能力を獲得してもらうことでモデルの妥当性を検討することを試みたが、人間の抽象化能力が予想以上に優れていて、学習しなくても情報の抽象化が観察されてしまい、結局、学習を通した抽象化能力の効果を明確に示すことができなかった。

実際の可動カメラを用いた矢印学習タスクでは、単に、ゴールに着いたときの報酬と、間違った方向に進んでしまった場合の罰に基づいて学習した結果、リカレントニューラルネット内に、矢印の向きを表現し、矢印が見えなくなった後もその情報を記憶するニューロンが見つかった。このことは、強化学習を通して、必要に応じて概念が自律的に形成されることを示唆するとともに、必要な情報を記憶することも強化学習によって学習できることを示した。４つのの矢印パターンを見せて学習し、学習に用いていない矢印パターンを見せたところ、ある程度の汎化能力が確認された。さらなる汎化能力の向上を期待し、画像のエッジ情報を追加することを試みたが、時間切れのため、教師あり学習のシミュレーションでの汎化能力の向上を示すに留まった。

一方、予測の学習については、図２のように、可変方向，可変速度で進む物体を、途中で見えなくなっても捕獲できるようになった。この際、捕獲タイミングと位置を予測する必要があるが、そのことを事前に与えることなく、学習を通してその必要性を自ら発見し、学習することができた。捕獲すべきタイミングは、物体の進行方向や速度によって連続的に変化する。２値の情報の記憶はポジティブフィードバックによって双安定状態を形成することで実現できることがわかっているが、連続的に変化する値(タイミング)をニューラルネットがどのように保持するかが注目された。内部を解析すると、物体が見えなくなる可能性のある直前の領域を通るのにかかる時間によってあるニューロンの値が変化し、その値によって別のニューロンの発火時間をずらし、それを複数のニューロンでリレーする形で捕獲タイミングの行動を起こすタイミングをずらしていることがわかった。つまり、ニューロン間のリレーで連続値情報を保持するという今までと違う記憶方式が学習によって獲得された。

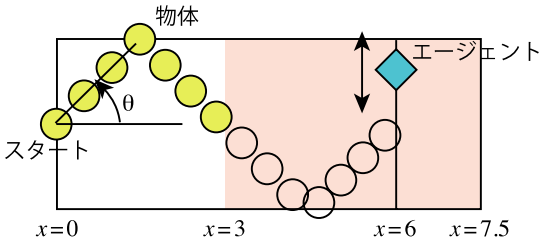

図２ 途中で見えなくなる物体の捕獲タスク

(3) 決定論的知的探索と時間的抽象化･好奇心

3x3 の限定された領域にランダムに壁が配置され、同じくランダムに配置された見えないゴールの探索タスクを学習することで、知的探索の学習を行なった。ほぼ最適に近い探索を学習によって獲得することができた。さらに、その際に、リカレントニューラルネットを使うことで、ゴール出現の可能性を予測しながら探索することができていることがわかった。また、図３のような分岐がある環境での探索タスクでは、図４のように、分岐の位置を記憶し、*t*=9 の行き止まりのとこ

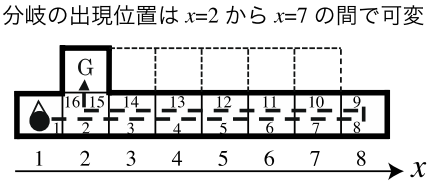

図３ 分岐がある環境の探索タスク

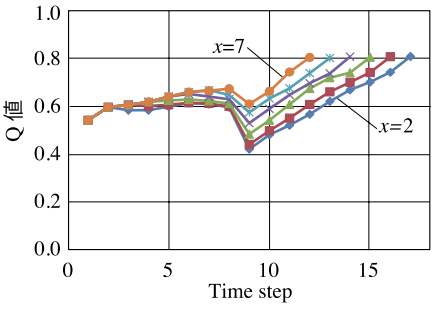

図４ 分岐位置による最大 Q 値の変化

ろまで来た際に、分岐の出現位置によって現在の状態の評価が異なることがわかった。また、この際にも、分岐の位置を記憶するために、中間層ニューロン間で値をリレーし、評価に反映させていることがわかった。

カメラを有する４足歩行の実移動ロボットを用いた実験では、相手のロボットにキスした時の報酬と見失った時の罰を与えるだけで、背景や照明条件が異なる 6000 個の画像信号から、相手のロボットの位置を把握し、80 から 90%の割合で相手のロボットにキスすることができるようになった。また、その前段階で行った首を振って相手のロボットの方を向くタスクでは、背景や照明条件によらないで相手のロボットを認識するニューロンが、強化学習をすることでニューラルネット内に発現したことを示した。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計９件）

① Kenta Goto and Katsunari Shibata: Acquisition of Deterministic Exploration and Purposive Memory through Reinforcement Learning with a Recurrent Neural Network, Proc. of SICE Annual Conf. 2010, 査読有, 2010, FB03-1.pdf
② Kenta Goto and Katsunari Shibata: Emergence of prediction by reinforcement learning using a recurrent neural network, Journal of Robotics, 査読有、Vol. 2010, 2010, Article ID 437654
③ Katsunari Shibata and Tomohiko Kawano: Acquisition of Flexible Image Recognition by Coupling of Reinforcement Learning and a Neural Network, SICE Journal of Control, Measurement, and System Integration (JCMSI), 査読有, Vol. 2, No. 2, 2009, pp. 122-129

〔学会発表〕（計１１件）

① 柴田克成, 沢津橋由人, 宇都宮浩樹:強化学習によるパターンの意味付けと記憶に基づく行動の獲得, 計測自動制御学会九州支部学術講演会, 2010 年 12 月 5 日, 宮崎大学
② 高津聡志, 柴田克成: 強化学習とリカレントネットを用いた並列で柔軟な学習制御システムの枠組み, SICE 九州支部学術講演会, 2010 年 12 月 4 日, 宮崎大学
③ Armad Afif bin Mohd Faudi, 柴田克成: 可動カメラを用いた Actor-Q 学習による能動認識の学習, 計測自動制御学会九州支部学術講演会, 2010 年 12 月 5 日, 宮崎大学

〔図書〕（計２件）

① Katsunari Shibata: Emergence of Intelligence through Reinforcement Learning with a Neural Network, Advances in Reinforcement Learning, Abdelhamid Mellouk (Ed.), InTech, 2011, pp.99-120
② 柴田克成,: ニューラルネットワーク, 「ロボット情報学ハンドブック」, ナノオプトニクスエナジー, 第 7.5 節, 2010, pp. 452-464

〔産業財産権〕
○出願状況（計０件）

○取得状況（計０件）

〔その他〕
ホームページ等

解説記事
柴田克成: 強化学習とニューラルネットによる知能創発, 計測と制御, Vol. 48, No. 1, 2009, pp. 106-111

６．研究組織
(1)研究代表者
柴田 克成（SHIBATA KATSUNARI）
大分大学・工学部・准教授
研究者番号：10260522

(2)研究分担者
なし

(3)連携研究者
なし