

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 12 日現在

機関番号：12701

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500275

研究課題名(和文) 強化学習を用いた進化論的生物型ロボットの行動形態獲得の研究

研究課題名(英文) Study on Movement-Pattern Acquisition of Biological Robot using Reinforcement Learning from Evolutional viewpoint

研究代表者

藪田 哲郎 (YABUTA, Tetsuro)

横浜国立大学・工学研究院・教授

研究者番号：30323926

交付決定額(研究期間全体)：(直接経費) 3,900,000円、(間接経費) 1,170,000円

研究成果の概要(和文)：強化学習を用いて人工生物型ロボット行動獲得を分析し、進化論的な観点からの考察を行おうとしたものである。昆虫、爬虫類、哺乳類と対称を変化させながら検討をすすめ、昆虫と爬虫類の比較では、脊椎がエネルギー効率の良い前進行動の獲得と密接な関係があることを明らかにし、哺乳類では行動の対称性がエネルギー効率の良い行動形態を少なくしていることを明らかにした。また、マルコフ性が保証されない大車輪ロボットを対象にし、確率的な挙動を可視化する手法を考案し、各種報酬と大車輪成功の関係を解明した。次に、人間が与える主観報酬は過去から未来への予測ができるので、現在のみに起因する客観情報よりも良い結果を与えることを示した。

研究成果の概要(英文)：This study clarifies movement-pattern acquisitions of biological robots using reinforcement learning method from evolutionary view point. By comparison between an insect and a reptile, a vertebral column gives essential factor to realize forward movements with small energy consumption. As for a mammal, symmetric condition of its movement-pattern decrease number of forward movement-patterns with small energy consumption. Next, visualization method of motion-pattern with probability clarifies relation between various rewards and success of giant swing for giant swing robot without Markov property. Finally, subjective rewards gives better performance than that of objective rewards because human being can estimate a sequence of movement-pattern from past to future although objective reward is obtained from a present condition.

研究分野：総合領域

科研費の分科・細目：情報学・感性情報学・ソフトコンピューティング

キーワード：強化学習 生物型ロボット 大車輪ロボット 報酬操作 ロボットハンド 人間計測

1. 研究開始当初の背景

赤ちゃんの「はいはい」行動を考えると、この動きは自分で獲得する行動であり、両親が「はいはい」の動きを教えているのではない。この問題は、現在の「学習」の枠組みでは、ロボットが教師なしで行動を獲得できるかどうかという問題設定であり、学習分野のみならず生物進化の問題の観点からも非常に興味ある問題である。しかし、ロボットの行動形態の獲得については教師ありの場合が殆どで、教師なしの場合についても外部からのクロック、さらには周期関数の情報を与えており、全く事前知識なしの条件下で教師なし学習を用いたロボットの行動形態獲得の研究は殆ど行われていない。

我々は、強化学習の評価関数が時系列の評価関数を持つ特徴を活用することで、教師なし学習でロボットの行動形態獲得を実現できることを明らかにした。具体的には、図1に示すイモムシ型ロボットを用いて、毎回、前進移動距離に比例した報酬を与えることで、図2に示す行動形態が発生することを明らかにした。この行動形態の発生機能は、ロボットが有限状態を遷移する時に、時系列報酬群(割引収益)を最大化するパターンとして学習される。従来の強化学習では、学習過程はブラックボックスで、最終的な学習結果にしか興味を持たれなかったが、この手法の面白さは、学習プロセスを図2に示すような行動形態の可視化が可能で、どのような力学的特性を活用して運動を実現しているかを解明することができる。

2. 研究の目的

この研究の背景にあるのは、人工生命の研究である。その研究は、コンピュータ内に人工生命を作り進化を評価するものであったが、マクロ的な生命の構造変化レベルが対象で、ダイナミクスを扱う行動形態のレベルにまでは発展しなかった。本研究では、上記手法を用いると設計者の意図を排除でき、全く同一条件で生物型ロボットの行動獲得とその力学的な解明ができるので、進化論の観点から生物の行動形態を解明することを試みる。また、生物では行動の対称性等が重要な要因であるが、生物型ロボットでは対称性の要因を外すことができるので、対称性の意味についても考察する。



図1 イモムシ型ロボット

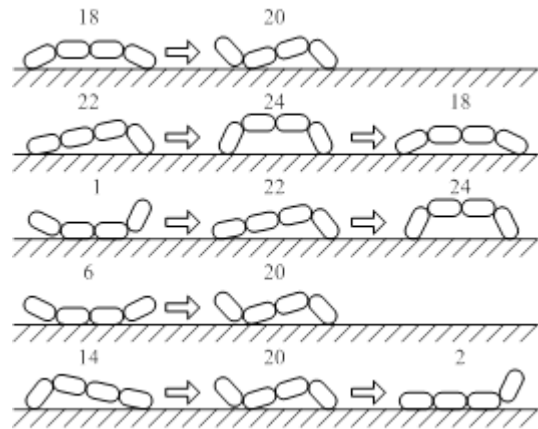


図2 イモムシ型ロボットの行動形態

また、強化学習を用いた大車輪ロボットの学習では、マルコフ性が保証できないので確率的な挙動が得られる。この扱いづらい非マルコフ性に起因する確率的な挙動の問題について解明を進める。

さらに、母親が子供に教える時は、「ほめる」等の主観的な報酬をもちいる。強化学習においても、この主観報酬は興味ある問題であるが、主観報酬と客観報酬の比較については検討例がないので、生物型ロボットを用いて、比較検討を進める。

上記の研究に関連する基本技術は、ロボット技術と人間計測技術である。上記の研究と平行して関連技術のポテンシャルを向上させるために、人間の柔らかさを実現するためのフィンガ・マニピュレータシステムのアドミッタンス制御とインピーダンス制御の研究を進める。さらに、人間計測については錯覚分野に関連するラバーハンド錯覚の研究と視覚と上肢運動の関係についても研究を行う。

3. 研究の方法

(1)強化学習を用いた生物型ロボットの行動獲得:

運動情報と消費エネルギーを報酬情報として、ダイナミックシミュレータを用いて、学習過程で得られる生物型ロボットの行動形態の強化学習を進める。また、この方法では姿勢情報を状態として用いることで、最終的に各種報酬に対する行動形態を求めることができる。

最初に、脊椎動物と無脊椎動物を対象に前進行動の行動形態を辿ってみることとした。具体的な対象としては、6足の昆虫タイプの生物型ロボット、4足の爬虫類タイプ、および4足で脊椎の自由度がない爬虫類タイプとした。報酬としては前進距離を用いる場合はその最大化、および前進距離と消費エネルギーの両者の場合は前進距離の最大化と消費エネルギーの最小化を図ることで、行動形態の比較を行った。

上記の昆虫、爬虫類の行動形態は静的な運動でダイナミクスの影響があまりでない。ダイナミクス効果についての面白い結果

としては、Hoyt らが馬の行動形態では、低速、中速、高速域で、それぞれに特有な行動モードが出現することを明らかにしている。本研究では、対象を哺乳類タイプとし、報酬として前進距離と消費エネルギーを用いて、このような行動モードが出現するかを検討した。

(2)非マルコフ性を有する大車輪ロボットの解析：

本研究の対象とする大車輪ロボットは位置制御タイプで、運動情報としては位置と速度しか観測できないので、マルコフ性が保証できない対象となる。このような対象に、強化学習結果を適用すると揺らぎの多い確率的な挙動となる。本研究では、このような非マルコフ性の揺らぎの多い強化学習の結果を評価するために、学習回数毎に確率的な指標を用いて可視化する方法を検討した。

(3)強化学習における主観報酬と客観報酬の比較：

イモムシ型ロボットを対象にして、移動距離を報酬とする客観報酬と、イモムシ型ロボットの姿勢から人間が報酬を与える主観報酬について前進距離の観点から比較検討を行った。

(4)フィンガ・アームロボットのインピーダンス制御とアドミッタンス制御の比較：

ロボットの柔らかさを実現する制御則に、インピーダンス制御とアドミッタンス制御が存在する。仮想バネを想定すると、前者は入力が位置で出力が力となり、後者は、その逆の制御方法となる。サーボ機構のギア等の影響で、両者の特性は良く分かっていない。本研究では、同一システムを用いて両者の基本的な傾向を解明した。

(5)ハプティックインタフェースを用いたラバーハンド錯覚：

ラバーハンド錯覚とは、自分の指で対象物を叩き、誰かに自分のもう一つ手の甲を同じようなタイミングで叩いて貰うと、自分で自分を叩いている錯覚に陥る。錯覚は人間の知覚メカニズムの解明、さらには錯覚を用いた新たなサービスにも芽になる可能性もある。他人を介在して実験をすると定量的な実験が困難なので、本研究ではハプティックインタフェースを用いて他人を介在せずに実験できるシステムを構築し、定量的な検討を進めた。

(6)受動型直線教示における再現描画時の視覚情報の影響：

バイラテラル型ハプティックインタフェースを用いると、視覚情報なしで線の描画を教示することができる。これを受動型教示と呼ぶが、その教示情報を再現するに際して、教示時に視覚情報がないのに、目の開閉が描

画長の影響する不思議な結果が得られた。この原因を解明するために、いろいろな実験条件下で再現実験を行い、その原因の解明した。

4. 研究成果

(1)強化学習を用いた生物型ロボットの行動形態の獲得：

前進距離を報酬として行動形態を学習させると、6足の昆虫タイプは昆虫と同様な行動パターンを、脊椎有の4足のヤモリタイプも図3に示すようにヤモリと同様な行動パターンを学習により獲得できた。さらに脊椎有の4足爬虫類タイプは脊椎なしのタイプに比較して、約20%のエネルギー改善が認められ、脊椎を前進行動に利用することでエネルギー効率の改善が図られることを明らかにした。

次に、前進距離を最大化しエネルギーを最小化するための報酬を用いて学習を進めると、エネルギー効率の最も良い行動形態を獲得することができた。

この手法を、より動的な馬型ロボットの行動形態の獲得を目指した。馬の歩行については、Hoyt らが低速、中速、高速で3つの歩行モードが存在することを示している。一方、馬型ロボットの学習結果では、低速から高速まで多数の行動形態が出現した。この学習では、行動形態に対称性、具体的には足の動きに対称性の条件が考慮されていないので、非対称性の行動形態も出現する。生物の行動形態には対称性が備わっているようであり、この対称性の差が、人工生物型ロボットと生物の行動形態の差として現れたものと思われる。

(2) 非マルコフ性を有する大車輪ロボットの解析：

非マルコフ性を有する大車輪ロボットの学習を行わせると、確率的な挙動の前回りおよび後回りの両者が混在する学習知識が得られる。この特徴を活かして、この学習知識をベースにして、前回りおよび後回りの確率が上昇できる学習知識ができるかどうかを検討した。これは人間と同様に、最初は大車輪達成の学習、それが達成できたら、前および後ろの方向を達成する2段階学習である。この大車輪ロボットでも、2段階学習が可能で、学習の選択性が実現できることを明らかにした。

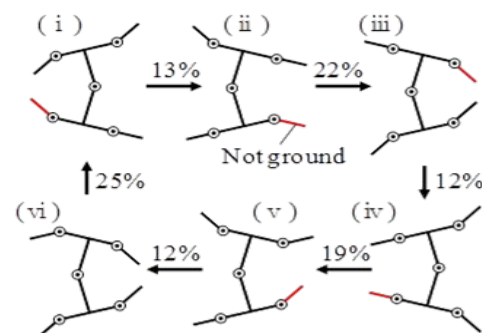


図3. 脊椎有のヤモリ型ロボットの行動形態

さらに、非マルコフ性を有する確率的な挙動を有する行動形態の解析方法としては、学習回数毎に前回の成功、後ろ回の成功、失敗の確率を計算することで、学習を進める毎の確率変化をグラフ上に可視化することで、報酬と成功要因の関係を明らかにすることができた。

(3) 強化学習における主観報酬と客観報酬の比較：

イモムシ型ロボットの前進行動獲得の学習において、センサから得られる前進移動量の客観報酬と、人間がロボットの姿勢の変化に与える5段階の主観報酬を用いて学習結果の行動形態の前進移動量を比較検討した。センサから得られる移動量はアナログ量で情報量が大きいため、この移動量も5段階のデジタル量に変換して、同一条件で比較を行った。客観情報のデジタル変換をどのように変化させても、このロボットの行動形態を熟知した人間の評価者が与える主観報酬が最大の移動量を与える行動を獲得した。この原因を分析すると、人間は良い行動形態の時系列パターンを知っており、そのような行動姿勢に高い評価を与えていた。

(4) フィンガ・アームロボットのインピーダンス制御とアドミッタンス制御の比較：

フィンガの対象物把持において、内力および外力に仮想インピーダンスを実現するために、内力外力分離型インピーダンス制御およびアドミッタンス制御の比較検討を行った。インピーダンス制御はギアの影響等で特性が劣化するが、アドミッタンス制御はフィードバック機能で特性改善が認められるが、一方、不安定性の問題が発生することを明らかにし、それぞれの特徴を解明することができた。

また、フィンガ部分の可操作度を最大に保ちながら、フィンガ部に柔らかさを与えるインピーダンス制御とアドミッタンス制御の比較検討を行った。通常用いられている位置/速度制御のマニピュレータを用いた場合には、同様な制御則を用いているアドミッタンス制御との相性の関連から良い結果が得られることを明らかにした。

最後に、ロボットに人間の作業スキルを組み込む観点から、マニピュレータにヨーヨー運動の実現を目指した。人間のヨーヨー運動を分析し、その特徴量をマニピュレータ制御の限界を考慮して操作量を構成し、目隠しヨーヨーを実現した。

(5) ハプティックインタフェースを用いたラバーハンド錯覚

ハプティックインタフェースを用いることで、ラバーハンド錯覚の定量的な考察ができる。ラバーハンドの最も重要な要因は、自分と他人の手が叩く同期のタイミングである。まず、この同期のずれがどの程度まで許

容できるのかを調べてみた。このタイミングの遅れは約200msまでで、それ以上の遅れが生じると錯覚効果が小さくなることを明らかにした。さらに、両手の水平方向および鉛直方向の空間位置を変化させて調べたところ、身体イメージが保てない距離および方向になると、錯覚現象が減じることを定量的に明らかにした。

(6) 受動型直線教示における再現描画時の視覚情報の影響：

バイラテラル型のハプティックインタフェースを用いると、視覚を用いずに教師の動きを追従する受動型の運動教示が可能である。教示された直線長をハプティックインタフェースで再現してもらおうと、視覚の有無で明確な有意差が認められた。

教示時に視覚を用いていないのに、再現時に視覚の有無の差が生じる不思議な結果である。ハプティックインタフェースを用いずに再現実験を行い、その原因解明を進めた。この現象は、再現時に新たな負荷が作用する時に発生し、視覚があると負荷作用時にも教示時の手先の速度プロファイルを保持しようとする効果に起因することを明らかにした。直線描画においての視覚の影響は、終端への位置制御機能と一般的には思われているが、速度制御機能も作用していることは興味深い結果である。

5. 主な発表論文等

〔雑誌論文〕(計 13 件)

河村拓実、松永和輝、藪田哲郎、閉眼受動型教示した線分の描画再生における機械的な負荷に対する有無が及ぼす影響、電子情報通信学会和文論文誌A、査読有、(掲載予定)

樋渡勇太郎、藪田哲郎、目隠しラバーハンド錯覚における触刺激間の距離および錯覚感度に与える影響、電子情報通信学会和文論文誌D、査読有、Vol.J97-D、No.5、2014、pp.1058-1062

山田大輔、佐野嘉則、堀良太、黄健、藪田哲郎、可操作度保持型マルチフィンガ・アームロボットの協調的アドミッタンス制御における動特性、日本機械学会論文集C、査読有、Vol.79、No.808、2013、pp.5010-5014

東浦拓也、松本悟、藪田哲郎、確率的ゆらぎを有する強化学習を用いた大車輪ロボットの行動獲得と報酬の関係について、日本機械学会論文集C、査読有、Vol.79、No.807、2013、pp.4335-4339

佐野嘉則、堀良太、藪田哲郎、内力外力インピーダンスモデルを用いた物体把持手法に関するインピーダンス制御とアドミッタンス制御の比較、日本機械学会論文集C、査読有、Vol.79、No.807、2013、pp.4330-4334

黒田将史、山科亮太、藪田哲郎、主観報酬を用いた強化学習における人間の教示

特性に関する考察、日本機械学会論文集 C、査読有、Vol.79、No.801、2013、pp.1770-1774

Daisuke YAMADA、Jiang HUANG、Tetsuro YABUTA、Comparison Between Admittance and Impedance Control of a Multi-Finger-Arm Using the Guaranteed Manipulability Method、Precision Instrument and Mechanology、査読有、Vol.2、Iss.1、2013、pp.85-93

山科亮太、黒田将史、藪田哲郎、主観報酬を用いた強化学習によるイモムシ型ロボットの行動獲得、日本機械学会論文集 C、査読有、Vol.79、No.798、2013、pp.366-371

佐野嘉則、堀貴之、早川拓人、藪田哲郎、人間技能の抽出・模擬によるフィンガ・アームロボットのヨーヨー操作（目隠しヨーヨーの実現）日本機械学会論文集 C、査読有、Vol.78、No.790、2012、pp.2074-2089

横山智宏、坂井直樹、藪田哲郎、強化学習を用いた大車輪ロボットの学習知識の選択可能性について、日本機械学会論文集 C、査読有、Vol.78、No.790、2012、pp.2090-2105

坂井直樹、豊田希、藪田哲郎、強化学習を用いた生物型ロボットの前進行動形態の獲得と機構がエネルギーに及ぼす影響、日本機械学会論文集 C、査読有、Vol.78、No.789、2012、pp.1886-1898

松永和輝、清政拓、山本健悟、豊田希、藪田哲郎、ハプティックインタフェースを用いた受動型直線描画教示における視覚情報の影響、電子情報通信学会和文論文誌 D、査読有、Vol.J95-D、No.2、2012、pp.347-350

堀貴之、佐野嘉則、黄健、豊田希、藪田哲郎、フィンガとマニピュレータの個別アドミッタンス制御におけるタスク実行時のフィンガ可操作度補償効果、日本機械学会論文集 C、査読有、Vol.77、No.776、2011、pp.196-207

〔学会発表〕(計 3 件)

Daisuke YAMADA、Yoshinori SANO、Ryota HORI、Jiang HUANG、Tetsuro YABUTA、Dynamic Movement by Admittance Control of a Multi-Finger-Arm Robot with Manipulability Control of Fingers、Proc. of SII、2013、pp.521-526

佐野嘉則、堀良太、堀貴之、藪田哲郎、内力外力インピーダンスモデルを用いたロボットフィンガによる実在物体把持、第 18 回ロボティクスシンポジウム予稿集、2013、pp.206-211

Ryota YAMASHINA、Masafumi KURODA、Tetsuro YABUTA、

Caterpillar Robot Locomotion Based on Q-Learning using Objective/Subjective Reward、Proc. of SII、2011、pp.1311-1316

〔その他〕

ホームページ等

<http://yabsv.jks.ynu.ac.jp>

6 . 研究組織

(1) 研究代表者

藪田 哲郎 (YABUTA TETSURO)

横浜国立大学・工学研究院・教授

研究者番号：30323926

(2) 研究分担者

豊田 希 (TOYOTA NOZOMI)

横浜国立大学・工学研究院・研究教員

研究者番号：60547222