

## 科学研究費助成事業 研究成果報告書

平成 26 年 6 月 25 日現在

機関番号：55201

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23560491

研究課題名(和文) 模倣学習と強化学習の融合による複数行動の効率的獲得

研究課題名(英文) Behaviors Acquisition for Multiple Tasks by Integration of Imitation Learning and Reinforcement Learning

研究代表者

堀内 匡 (HORIUCHI, Tadashi)

松江工業高等専門学校・電子制御工学科・准教授

研究者番号：50294129

交付決定額(研究期間全体)：(直接経費) 4,000,000円、(間接経費) 1,200,000円

研究成果の概要(和文)：本研究では、実機ロボットを対象として、模倣学習と強化学習を融合した枠組みにより複数タスクに対する行動をロボット自身が効率的に獲得する方法について検討した。まず、単一の行動タスクに対する研究として、1) 尺取虫型ロボットに対する模倣学習と行動学習の融合、2) 四脚ロボットに対するCPGと強化学習の融合の二つを実現した。さらに、複数の行動タスクに対する研究として、3) 二脚駆動ロボットに対する複数タスクの模倣学習、4) 四脚ロボットに対するCPGと強化学習の融合による複数タスクの獲得、の二つを実現した。

研究成果の概要(英文)：In this research, we realized the effective acquisition of multiple behaviors for real robots by integration of imitation learning and reinforcement learning. First of all, we realized the behavior learning for single task such as 1) integration of imitation learning and behavior learning for the measuring worm robots and 2) integration of CPG and reinforcement learning for the four-legged robots. Moreover, we realized the behavior learning for multiple tasks such as 3) imitation learning of multiple behaviors for the two-armed gating robots and 4) effective acquisition of multiple behaviors by integration of CPG and reinforcement learning for the four-legged robots.

研究分野：工学

科研費の分科・細目：電気電子工学・システム工学

キーワード：知能ロボティクス 模倣学習 強化学習 行動獲得 複数タスク

## 1. 研究開始当初の背景

強化学習は、環境と相互作用し、自分自身で試行錯誤を繰り返すことでタスク達成を目指す学習方法であり、多くの研究がなされている。しかし、問題が複雑で大規模になると膨大な試行錯誤の回数を必要とする欠点があり、ロボットなどでの実用化はまだ少ない。一方、他者を真似ることによりタスク達成を目指す模倣学習の研究も盛んになされている。模倣は、行動観察に基づく学習であり、強化学習より効率的であるとも言われているが、教示者の動きがタスク達成に対して最適化された動きとは限らないなどの問題点がある。

そこで、模倣学習と強化学習を組み合わせた枠組みが必須であると考えられる。つまり、模倣学習により他者の行動を参考にして大まかな行動知識を獲得するとともに、試行錯誤に基づく強化学習を継続的に行い、より洗練された行動に変化させてゆく枠組みである。本研究ではこの枠組みを単一の学習器として構成するのではなく、複数のタスクを達成する多様な行動を獲得する自己組織化型の学習システムとして構成することを検討する。

## 2. 研究の目的

本研究では、実ロボットを対象として、模倣学習と強化学習を融合した枠組みによる複数行動の効率的獲得の実現を目指している。実ロボットとして、LEGO社のMindstorms NXTやROBOTIS社のBioloidで組み立てた数種類のロボットを対象とする。具体的には以下の二つを研究の目的とする。

第一の目的として、単一の行動タスクに対する模倣学習と強化学習の融合による効率的な行動獲得を実現する。さらに、第二目的として、複数の行動タスクを考え、それらのタスクに対する模倣学習と強化学習の融合による多様な行動の獲得を実現する。

## 3. 研究の方法

まず、単一の行動タスクに対する研究として、以下の方法により研究を実施した。

### (1) 尺取虫型ロボットに対する模倣学習と行動学習の融合の実現

まず、ROBOTIS社のBioloidを用いて2台の尺取虫型ロボットを組み立てた。その尺取虫型ロボットを対象として、模倣学習と行動学習の融合を実現した。行動学習として、強化学習に近い枠組みである同時摂動最適化法を用いて、周期運動のパラメータを学習した。

### (2) 四脚ロボットに対するCPGと強化学習の融合の実現

まず、ROBOTIS社のBioloidで四脚ロボットを組み立てた。その四脚ロボットを対象として、CPG(中枢パターン生成器)と強化学習の融合を実現した。四脚ロボットのウォークと呼ばれる歩行パターンを観測し、その歩行パターンを模倣するCPGを構成した。そのCPGのパラメータを強化学習により最適化することにより、四脚ロボットが環境に応じて滑らかに前進する行動を獲得することを実現した。

次に、複数の行動タスクに対する研究として、以下の方法により研究を実施した。

### (3) 二脚駆動ロボットに対する複数タスクの模倣学習

まず、LEGO社のMindstorms NXTを用いて二脚で漕ぐように動く二脚駆動ロボットを組み立てた。その二脚駆動ロボットを対象として、階層型ニューラルネットワークを用いて、複数の行動タスクを模倣することを実現した。二脚駆動ロボットの複数の行動タスクとして、直進・右折・左折などの行動パターンを考えた。

### (4) 四脚ロボットに対するCPGと強化学習の融合による複数タスクの獲得

単一の行動タスクに対する研究(2)で組み立てた四脚ロボットを対象として、直進するだけでなく右折や左折も滑らかにを行い、与えられた目標方向へ進む行動パターンの獲得および与えられた目標地点に到達して停止する行動パターンの獲得を実現した。研究(2)と同じように、四脚ロボットのウォークと呼ばれる歩行パターンを観測し、その歩行パターンを模倣するCPGを構成した。そのCPGのパラメータのうち、左前脚と右前脚のモータの振幅を強化学習により最適化した。それにより、滑らかに直進・右折・左折する行動を獲得でき、目標方向に向きを変えて前進する行動の獲得および目標地点に到達して停止する行動の獲得を実現した。

## 4. 研究成果

### (1) 尺取虫型ロボットに対する模倣学習と行動学習の融合の実現

ROBOTIS社のBioloidで組み立てた性能の異なる2台の尺取虫型ロボット(図1参照)を対象として、模倣学習と行動学習の融合を実現した。2台のロボットのうち、一方を教示者ロボット(最大トルク大)、もう一方を学習者ロボット(最大トルク小)とする。各ロボットのモータのトルクが異なるため、学習者ロボットは教示者ロボットと同じ行動をしても前進することはできない。観察した行動をそのまま再現するだけでなく、模倣により獲得した行動パターンを利用して自身に適した行動を学習する必要がある。

具体的には、まず教示者ロボットの前進行動を Web カメラで観測し、その前進行動パターンを周期運動として捉え、1 周期の運動を台形モデルで近似した。尺取虫型ロボットを対象とした実験において、学習者ロボットに効率的に前進行動を学習させるため、行動学習の手法として同時摂動最適化法を用いて、周期波形から推定したパラメータを変更し、学習者ロボットの前進に適した行動を獲得することを実現した。

実験結果の一例を図 2 に示す。図 2 は、模倣により獲得した行動パターンを初期点として、同時摂動最適化法による行動学習の結果を示している。

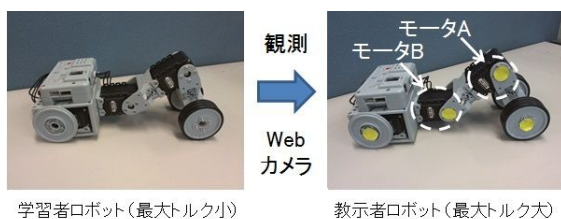


図 1 尺取虫型ロボット

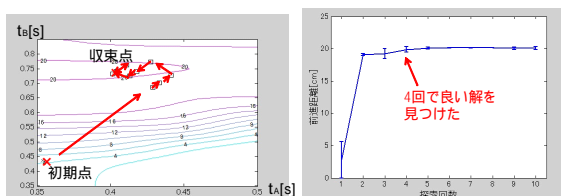


図 2 実験結果

## (2) 四脚ロボットに対する CPG と強化学習の融合の実現

ROBOTIS 社の Bioloid で組み立てた四脚ロボット (図 3 参照) を対象として、CPG (中枢パターン生成器) と強化学習の融合を実現した。四脚ロボットのウォークと呼ばれる歩行パターンを観測し、その歩行パターンを模倣する CPG を構成した。その CPG モデルのパラメータを強化学習により最適化することで、四脚ロボットが環境に応じて滑らかに前進する行動を獲得することを実現した。

具体的には、まず三角関数を用いた CPG モデルにより、ウォークと呼ばれる歩行パターンを模倣した。その CPG モデルのパラメータを強化学習の代表的な手法である Q 学習により最適化することで、実機の四脚ロボットが目標方向に滑らかに前進する行動パターンを獲得することに成功した。従来の多くの研究では、ロボットの各モータの各時刻における目標角度そのものを学習対象としていたが、この研究では CPG のパラメータを強化学習により学習している。これにより生成した波形に基づいてモータを連続的に動かすことで、四脚ロボットの滑らかな行動パターンを実現している。



図 3 四脚ロボット

## (3) 二脚駆動ロボットに対する複数タスクの模倣学習

LEGO 社の Mindstorms NXT で組み立てた二脚駆動ロボット (図 4 参照) を対象として、階層型ニューラルネットワークを用いて、複数の行動タスクを模倣することを実現した。二脚駆動ロボットに対する複数の行動タスクとして、直進・右折・左折などの行動パターンを考えた。

まず、教示者ロボットのそれぞれの動きをモーションキャプチャシステムにより観測した。次に、階層型ニューラルネットワークを用いて、ロボットのダイナミクスを表す逆モデルを構築し、教示者の行動の推定に必要なモータ出力を求めた。さらに、教示者の行動を複数の順モデルでモデル化し、複数の行動パターンを模倣・生成することを実現した。図 5 は実験結果の一例を示す。図 5 は、直進・右折・左折の行動パターンを学習者ロボットが模倣・生成した結果を示している。

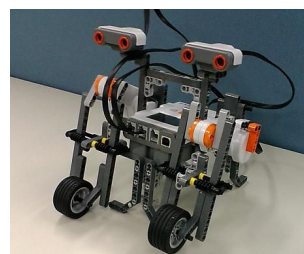


図 4 二脚駆動ロボット

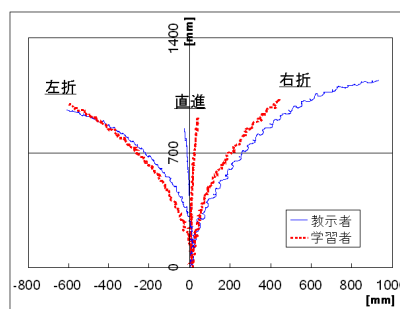


図 5 獲得した複数の行動パターン例

## (4) 四脚ロボットに対する CPG と強化学習の融合による複数タスクの獲得

単一タスクに対する研究(2)では、四脚ロボットが滑らかに前進する行動パターンを

獲得した。その成果を踏まえて、直進するだけでなく右折や左折も滑らかに行い、与えられた目標方向へ進む行動パターンの獲得および与えられた目標地点に到達して停止する行動パターンの獲得を実現した。

具体的には、図2に示した四脚ロボットを対象として、三角関数を用いたCPGモデルにより、ウォークと呼ばれる歩行パターンを模倣した。CPGのパラメータのうち、左前脚と右前脚のモータの振幅を強化学習により最適化することで、目標方向に向きを変えて大きく前進する行動の獲得および目標地点に到達して停止する行動の獲得を実現した。強化学習の手法としてはQ学習を用いた。

目標方向へ進む行動の獲得では、四脚ロボットの向きと目標方向の角度を状態変数、左前脚と右前脚のモータの振幅に対応するCPGパラメータを行動変数として、離散的な状態行動空間を構成した。目標地点に到達して停止する行動の獲得では、四脚ロボットの向きと目標地点の角度に加えて目標地点までの距離も状態変数とした。

実験結果の一例を図6～図8に示す。図6は学習初期の四脚ロボットの移動軌跡を示し、学習初期では目標地点に向かっていないことがわかる。図7は学習後の四脚ロボットの移動軌跡を示し、目標地点に向かって向きを変えて進む行動の獲得に成功していることがわかる。図8は目標方向への移動距離の推移を示し、着実に目標方向への移動距離が伸びていることがわかる。

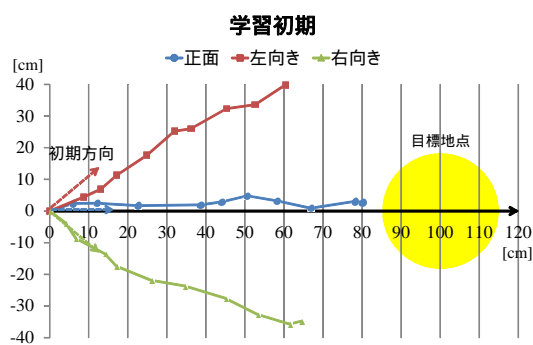


図6 学習初期の四脚ロボットの移動軌跡

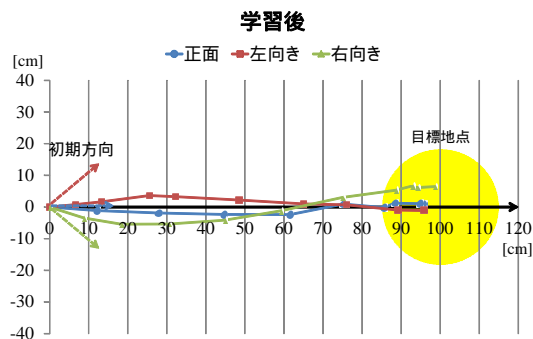


図7 学習後の四脚ロボットの移動軌跡

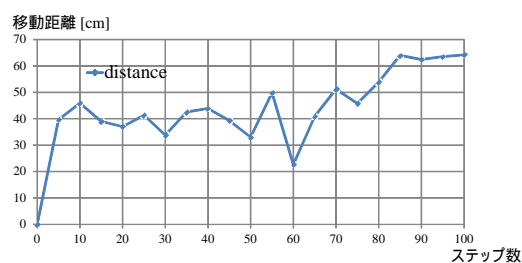


図8 実験結果(移動距離の推移)

## 5. 主な発表論文等

〔雑誌論文〕(計6件)

K. Ishihara, Y. Ohmura and T. Horiuchi, Behavior Acquisition of Six-Legged Robot by Reinforcement Learning using Value Iteration, ICIC Express Letters, 査読有, Vol.7, pp.1385-1391, 2013

石原弘二, 大村佑貴, 堀内 匡, モデルベース強化学習を用いた六脚ロボットの前進行動獲得, 計測自動制御学会論文集, 査読有, Vol.49, pp.408-410, 2013

K. Ishihara, T. Nagase and T. Horiuchi, Behavior Acquisition of Gating Robot by Reinforcement Learning using NGnet, ICIC Express Letters, 査読有, Vol.6, pp.1083-1089, 2012

H. Uchida and T. Horiuchi, A Study on Acquisition of Walk Movement by Imitation Learning for Real Robot, ICIC Express Letters, Part B: Applications, 査読有, Vol.2, pp.577-582, 2011

〔学会発表〕(計35件)

岸本良一, 石倉裕貴, 堀内 匡, CPGと強化学習を用いた四脚ロボットの目標到達行動の獲得, 電気学会システム研究会, ST-14-016, 2014年3月, 千葉県千葉市

石倉裕貴, 堀内 匡, CPGと強化学習を用いた六脚ロボットの前進行動獲得の試み, 第29回ファジィシステムシンポジウム, 2013年9月, 大阪府枚方市

堀内 匡, 金津達也, 石倉裕貴, 石原弘二, CPGを用いた多脚ロボットの歩行パターン獲得に関する検討, 平成25年電気学会電子・情報・システム部門大会, 2013年9月, 北海道北見市

畑山綱毅, 堀内 匡, 模倣学習による尺取虫型ロボットの行動獲得, 第21回計測自動制御学会中国支部学術講演会, 2012年11月, 広島県広島市

宇田川大二郎, 堀内 匡, 運動学習プリミティブを用いたロボットアームの動作模倣, 第21回計測自動制御学会中国支部学術講演会, 2012年11月, 広島県広島市

石原弘二, 石倉裕貴, 大村佑貴, 堀内 匡, 価値反復による強化学習を用いた六脚ロボットの前進行動獲得, 第28回ファジィ

システムシンポジウム，2012年9月，  
愛知県名古屋市

加藤 聡，堀内 匡，自己組織化マップを用いたクラスタリングにおける情報量規  
準によるクラスタ数の推定，第28回ファ  
ジィシステムシンポジウム，2012年9  
月，愛知県名古屋市

内田大貴，宇田川大二郎，堀内 匡，模倣  
学習を用いた尺取虫型ロボットの複数行  
動獲得，第54回自動制御連合講演会，  
2011年11月，愛知県豊橋市

宇田川大二郎，内田大貴，堀内 匡，模倣  
学習による二脚駆動ロボットの行動獲得，  
第27回ファジィシステムシンポジウム，  
2011年9月，福井県福井市

加藤 聡，GPUによるSOM学習アルゴリズムの高速化に関する研究，第27回ファジ  
ィシステムシンポジウム，2011年9月，  
福井県福井市

石原弘二，永瀬孝洋，堀内 匡，NGnet を  
用いた強化学習による歩容ロボットの前進  
行動獲得，平成23年電気学会電子・情報・システム部門大会，2011年9月，富  
山県富山市

畑山綱毅，内田大貴，堀内 匡，Web カメ  
ラを用いた模倣学習による尺取虫型ロボ  
ットの行動獲得，平成23年電気学会電  
子・情報・システム部門大会，2011年9  
月，富山県富山市

## 6. 研究組織

### (1) 研究代表者

堀内 匡 (HORIUCHI, Tadashi)  
松江工業高等専門学校・電子制御工学科・  
准教授  
研究者番号：50294129

### (2) 研究分担者

加藤 聡 (KATO, Satoru)  
松江工業高等専門学校・情報工学科・  
准教授  
研究者番号：40342547