1E1-02

強化学習を利用した二足歩行ロボットのための学習コントローラーの設計

Design of a controller for a two-legged robot using the framework of reinforcement learning

平石 広典[*1] 石黒 駿太郎[*1]
Hironori Hiraishi Syuntaro Ishiguro

[*1] 秋田工業高等専門学校 電気情報工学科
Department of Electrical & Computer Engineering, Akita National College of Technology

We have designed a controller for a two-legged robot, which allows a robot to learn target behaviors from human operations. The controller records sequences of user operation. And the controller does not make a robot only execute recorded behaviors again, but also a robot can execute more effective behaviors using the framework of reinforcement learning. So, the controller learns behaviors from trial and error of user operation and achieves more effective behaviors for a two-legged robot.

1. はじめに

現代，様々な種類のロボットが登場している．人間のパートナーとして日常生活の家事などの負担を軽減し，コミュニケーションの相手としてロボットが活動することを目指して作られている．しかし，実際の生活に取り入れるのはまだ難しく，エンターテイメントや，ある特定の目的で限られた場所にのみ利用されているのが現状である．ロボットを日常生活に取り入れることが難しい理由としては，コストが高いことや，自律性と操作性の両立に高い技術が必要になることがあげられる．

本研究ではこの自律性と操作性の問題について着目した．

- 自律性
  ロボットが自ら考えて行動することで，ある特定のタスクを自動的に達成する能力だが，これを実現するためには，高度なセンサ技術やプログラミング技術が必要となる．
- 操作性
  ロボットを人間が直接操作することで多様なタスクへの対応が可能になるが，ユーザの精密な操作が必要となる

これら両方の特性を実現するために，強化学習の枠組みを利用する．本研究では，二足歩行ロボットを対象とし，人間の試行錯誤の操作から，目的の行動を学習することのできる学習コントローラーの設計する．これによりロボットは半自動化され，ユーザの操作の負担を軽減し，的確でスムーズな動作を実現するものである[田淵 2006]．

2. 二足歩行ロボット

二足歩行ロボットは，タイヤを持つロボットとは異なり，歩くなどの移動動作の区切りがはっきりしており，一つ一つの動作が明確である．そのため，ある目的の行動を行うための動作系列が明確であり，学習する際に，一歩前進，一歩後退など動作の選択がわかりやすいといったメリットがある．

本研究では図1に示している，タカラトミー社製の二足歩行ロボット「i-SOBOT」を使用した．このロボットは全長が 16.5cm であり，17 自由度をもっている．またロボットを安定して動作させるためにジャイロセンサーが搭載されている．

連絡先：平石広典，秋田工業高等専門学校，秋田県秋田市飯島文京町１番１号，電話&Fax 番号 018-847-6042，hiraishi@akita-nct.jp，http://akita-nct.jp/hiraishi/

図 1：二足歩行ロボット「i-SOBOT」

また，パソコンから赤外線リモコンの信号を発信可能な BUFFALO 社のパソコン用学習リモコン「PC-OP-RS1」を使用することで，パソコン上からロボットの操作コマンドを送信することが可能であり[中川 2008]，プログラムを通して，ユーザ操作の記録と学習された行動の実行が可能である．

3. 強化学習の枠組みの利用

強化学習では，エージェントがある状態において取るべき動作を選択し，その動作によって目的が達成された場合に，環境から報酬を得る．そして，一連の動作を通じて報酬が最も多く得られるような行動を学習するといった枠組みである[三上 2000]．環境から得られる報酬は，目的を達成した直前の動作に対して与えられ，それ以前の動作には，結果として減衰した値が与えられるのが一般的である．これによって，より目的を達成しやすい動作を選択させることができる．

本研究では，このような強化学習の枠組みを利用し，ゴールに到達した直前の動作に 1.0 の報酬を与え，それ以前の動作は 0.9，0.8，0.7 というように報酬の値を減衰して与えることとした．図 2 には本研究における報酬の分配の例を示した．中央に示しているルートのスタート直後の動作は，ゴールから報酬を減衰させていった時の二通りの動作の直前の動作であり，左側のルートの報酬 0.8 と右側のルートの報酬 0.7 を足し合わせた 1.5 の報酬が与えられている．そのため，図 2 の例では**[スタート]→[1.5]→[0.9]→[1.0]→[ゴール]**といった行動を選択し実行す

るようになる．したがって，ロボットは報酬の高い動作を選択すれば，より目的を到達できる可能性の高い行動を行うことができる．

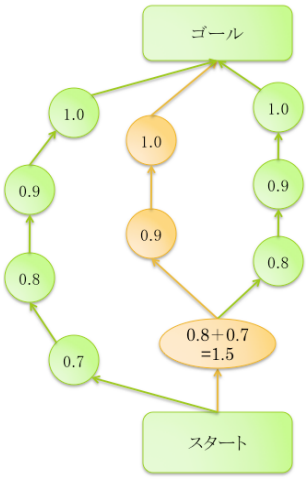

図 2:ロボットの動作における報酬の分配

4. 学習コントローラー

図3は設計した学習コントローラーを示した．左側の矢印ボタンによってロボットの基本的な移動が可能であり，左下のアクションを選択して，ロボットがもともと持っているアクションを実行することができる．右側には学習した行動が表示されており，行動を選択して実行ボタンを押すことで，学習した行動を実行することができる．ロボットに学習させるためには，**[学習行動の登録]→[学習開始]→[コントローラーによる操作]→[学習終了]**といった流れである．

図3:学習コントローラー

5. 実験結果

本研究では，以下の3つのタスクで実験を行った．

- タスク 1　前方の障害物をよける短いコース
- タスク 2　前方に置いてある物体をゴールに運ぶコース
- タスク 3　距離が長く，曲がることが必要となるコース

それぞれのタスクについて5回ずつ学習回数を増やしていく．そして，学習の完了後に，そのタスクを実行し，確実にゴールに到達した場合を成功として判断する．10 回の実験を行い，そのときのタスクの成功率を調べた．表1にはそれぞれのタスクについて，学習回数を増やした場合の成功率の変化をまとめたものである．どのタスクにおいても学習回数を重ねるにつれて成功率が上昇していることがわかる．

さらに，タスク 3 においては，手動で作成したプログラム(開発調整に数日程度)で実験した結果，成功率は 90%であった．強化学習を用いたものは 15 回の学習で 80%の成功率であり，学習にかかった時間はおよそ 30 分ほどであった．これより本コントローラーを利用することでより短い時間で目的の動作を実現することが可能である．

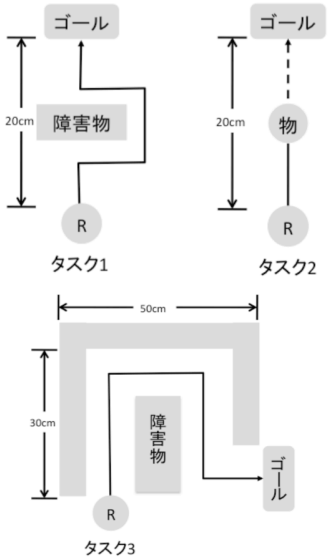

図 3:実験タスク

表1:実験回数とタスク成功率(%)

実験回数	5	10	15
タスク 1	60	50	70
タスク 2	70	70	80
タスク 3	70	80	80

6. まとめと今後の課題

本研究では，強化学習の枠組みを利用した二足歩行ロボットのための学習コントローラーの設計を行った．実験より，ユーザの試行錯誤の操作より，目的の行動を学習することが可能であることを明らかにした．

実験結果で示したタスク 3 において，プログラミングした場合とわずかな差が出たのは，ロボットの連続する動作の間に原因があると考えられる．プログラミングしたロボットの動作には成功率を上げるために，動作後にわずかな停止時間を挟んだケースがある．しかしながら，学習を用いた行動では，どの程度の停止を入れるかの判断は難しく，よりスムーズな行動を実現するために，このような停止を入れていない．これによって安定性に差が出たと考えられ，停止をどのように判断するかが今後の課題である．

参考文献(論文誌と同じスタイルを推奨)

[田淵 2006] 田淵一真，谷口忠大，椹木哲夫: “模倣学習と強化学習の調和による効率的行動獲得”，人工知能学会全国大会(第 20 回)論文集, 2006.

[中川 2008] 中川信行:PLUS ROBOT vol.1，株式会社 毎日コミュニケーションズ, 2008.

[三上 2000] 三上貞芳，皆川雅章:強化学習，森北出版株式会社，第1版, 2000.