

A2-7 倒立振り子型自律移動ロボットのための方策勾配法を基づく動的動作学習

福井大学 工学部 知能システム工学科 進化ロボット研究室
野々下 博昭 (指導教員：高橋 泰岳、前田 陽一郎)

1. 緒言

人間との共生環境下において活動する移動ロボットの開発が進んでいる [1]。従来の移動ロボットの多くは地面への投影面積を大きくし、静的安定性を保障している。しかし、人間との共生環境下において人間と同程度の動作を実現するには、この静的安定性を保障した移動ロボットは地面への投影面積が大きくなりすぎるため向いていない。そのため、近年では人間との共生環境下において活動するヒューマノイド型ロボットの開発が進んでおり、転倒を防ぐための安定化技術が重要な課題の一つとなっている [2]。その中でも、倒立振り子を規範とした2足歩行による研究が行われており、車輪型機構においても倒立二輪型移動機構の研究が盛んに行われている [3]。

従来の倒立二輪型移動ロボットの研究では、安定した移動や倒立を問題にしたものが多い。しかし、人間との共生環境下において人間が見せるようなダイナミックな動作を実現するためには、倒立の安定性をわざと崩す必要がある。更に、一時的に不安定な状態を通して最終的には転倒せずに一連の動作を遂行しなくてはならない。このためには、従来のようにある一定の目標姿勢と目標速度を実現する制御以外に、ダイナミックで安定した動作を実現するための目標姿勢の軌道を生成する必要がある。

そこで本研究では、制御を加えなければ不安定な状態となる倒立振り子型移動ロボットを対象として、身体ダイナミクスを活用した蹴球動作の目標姿勢を強化学習の一つである方策勾配法によって学習させる。また、提案手法の有効性を確認するために倒立振り子型ロボットのシミュレータを開発した。

2. 提案手法

蹴球動作を実現するために2層の制御器を用意する。

2.1 蹴球動作の表現

蹴球動作は図1のようにモーションを2つに分けて表現し、ロボットの胴体を前後に傾けてロボットの胴体にボールを当てるように行う。

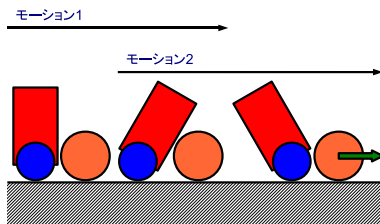


図1 蹴球動作の概略

2.2 目標姿勢生成器 (上位層)

上位層では、蹴球動作を表現する2つのモーションのパラメータを生成する。このパラメータを Kohl and Stone によって定式化された強化学習の1つである方策勾配法 [4] によって学習させる。この時、モーション1つにつき生成されるパラメータは表1の通りである。モーションは2つあるため、それぞれのパラメータの総計14個が学習パラメータとなる。

動作時間	t
目標胴体角	θ_d
目標車輪角速度	$\dot{\phi}_d$
胴体角誤差ゲイン	k_1
胴体角速度ゲイン	k_2
車輪角速度ゲイン	k_3
車輪角速度誤差ゲイン	k_4

表1 モーションパラメータ

2.3 姿勢制御器 (下位層)

上位層によって生成されたパラメータはモーションを実行する時間 t と蹴球動作を実現するのに必要な式 (1) のパラメータによって構成される。下位層では、これらの総計14個のパラメータを受けて蹴球動作を実現する制御を行う。

$$T = k_1(\theta - \theta_d) + k_2\dot{\theta} + k_3\dot{\phi} + k_4 \sum (\dot{\phi} - \dot{\phi}_d) \quad (1)$$

この時、 T は車輪トルク、 θ は胴体角、 θ_d は目標胴体角、 $\dot{\phi}$ は車輪角速度、 $\dot{\phi}_d$ は目標車輪角速度、 k_1 は胴体角誤差ゲイン、 k_2 は胴体角速度ゲイン、 k_3 は車輪角速度ゲイン、 k_4 は車輪角速度誤差ゲインである。

2.4 評価関数

蹴球動作の学習の評価は、ロボット本体がなるべく初期位置から移動しないことが望ましく、ボールを遠くまで蹴り出せる方が良いとする。また蹴り出した後に転倒する事も考えられるため転倒時はペナリティを与える。これを踏まえて評価関数を式 (2) のように設定する。

$$E = \begin{cases} \frac{w_1}{l_b} + w_2 v_b - w_f & (\text{転倒した場合}) \\ \frac{w_1}{l_b} + w_2 v_b & (\text{転倒しなかった場合}) \end{cases} \quad (2)$$

この時、 l_b はロボット本体の積算移動距離、 v_b はボールの初速度、 w_1 、 w_2 、 w_3 はそれぞれ、本体の積算移動距離、ボールの初速度、転倒時の重み係数である。

3. シミュレーション

ロボットのモデル図とシミュレータ画面をそれぞれ図2、図3に示す。評価の重みによって学習の進み方が変わることが考えられるため、ボールの初速度の重みを1000、転倒時の重み(ペナリティ)を-100で固定し、本体の積算移動距離の重みについては1、10、100、1000と変えて実験を行い、それぞれの学習の進み方がどのように変化するかを確認した。1学習の試行回数を50回とし、これを1000回学習することにした。

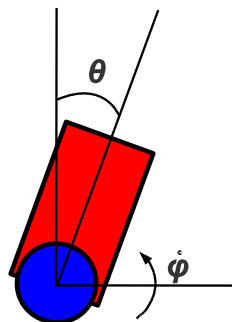


図2 モデル図

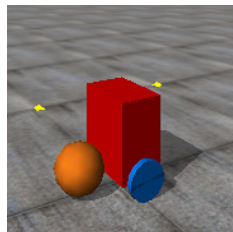


図3 シミュレータ画面

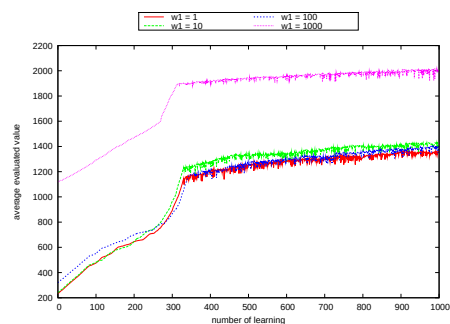
図4(a)、(b)、(c)、(d)、(e)より0から100回の間、100から300回の間、300から330回の間、そして330回以降の学習で学習傾向が変化していることが分かる。この変化は、シミュレータ上のロボットが学習におけるコツのようなものを習得し、その後の収束傾向の学習に向かったのではないかと考えられる。

4. 結言

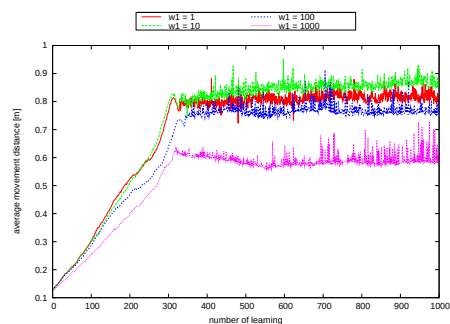
本研究では、倒立振り子型ロボットを対象として蹴球動作の目標姿勢を方策勾配法によって獲得させ、その挙動を動力学シミュレータによって調べた。今後の課題としては、今回作成したシミュレータの環境と実機ロボットにおける環境とでは大きな差異があるため、実機ロボットにおいても本手法が実現できるようにより細かい分析を行う必要がある。

参考文献

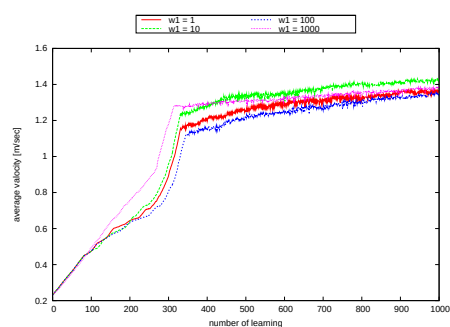
- [1] 菅野 重樹: “人間共生型のロボット,” 電子情報通信学会誌, Vol.88, No.5, pp.323-328 (2005)
- [2] 亀田 幸季, 関口 暁宣, 妻木 勇一, Dragomir N. NENCHEV: “特異点適合法によるヒューマノイドロボットの静歩行制御,” 日本機械学会論文集, C編, Vol.73, No.727, pp.790-795 (2007)
- [3] 中代 重幸, 古屋 麻衣子: “倒立振子を応用した車輪型移動ロボットに関する研究,” 日本機械学会関東支部総会講演会講演論文集, pp.383-384 (2002)
- [4] Nate Kohl, Peter Stone: “Policy gradient reinforcement learning for fast quadrupedal locomotion,” *Proceedings of the IEEE International Conference on Robotics and Automation 2004*, pp.2619-2624 (2004)



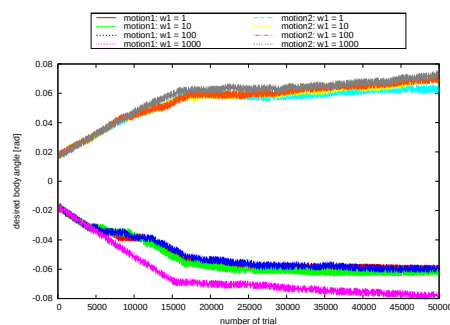
(a) 平均評価値



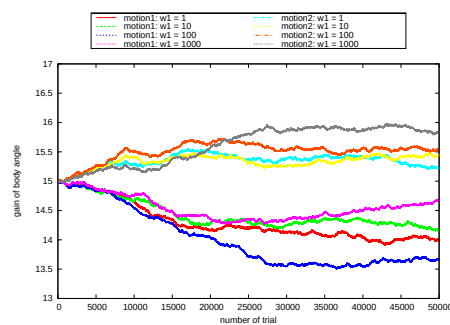
(b) 本体の平均移動距離



(c) ボールの平均初速度



(d) 目標胴体角



(e) 目標胴体角誤差ゲイン

図4 学習結果