

A3-2 倒立二輪型移動ロボットによる人間動作の模倣学習に関する研究

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

1. 緒言

近年、人間共生環境下における人間共生ロボットの研究が盛んに行われている [1]。人間共生ロボットの多くは、人間の行動空間において人間の行動を妨げることなく生活を支援し、人間の生活環境をより快適にすることが求められる。しかし我々の生活環境におけるロボットに要望する行為は日々異なるため、事前に全ての行為を設計することは困難である。そこで人間の動作データを基にロボットが模倣学習する研究が盛んに行われている [2]。しかし、人間とロボットでは身体構造が大きく異なるため、人間から得た動作データをそのまま模倣に用いるのは困難である。ロボットには人間の動作をロボット自身に適した動作として認識させる必要がある。

サッカー競技のような人間と対等にインタラクションするロボットを製作する場合、人間と同程度の背の高さを必要とするが、3輪以上の車輪型移動機構で安定性を求める場合、地面への投影面積が大きくなり、人間の日常生活環境下での稼働が難しくなる。このため、人間共生ロボットの移動機構には、地面への投影面積を小さくし、人間と同程度の大きさ、速さで動作が行える移動機構が必要となる。これらを実現する移動機構として倒立二輪型移動機構が研究されている。倒立二輪型移動ロボットは、一時的に不安定な状態を通すことで全身のダイナミクスを利用した動作が行えると考えられる。そこで本研究では、人間共生環境下に適する移動機構をもつ倒立二輪型移動ロボットを対象として人間が行う蹴球動作を観測させ、行動学習を通し、最適な目標軌道を生成する模倣学習を行う。この模倣学習のフローを図1に示す。

本報告では、人間の動作認識から得た動作データを経路点データによって再現し、ロボットが動作を行う際の目標軌道を生成する。いくつかの経路点データによる人間の動作データの再現はスプライン補間法によって行い、動作認識によって得た動作データをロボットに適した滑らかなスプライン曲線に近似する。このスプライン曲線を目標軌道とし、目標軌道を生成する経路点データを学習パラメータとすることで、少ない学習コストで行動学習を行うことができる。しかし、人間と倒立二輪型移動ロボットでは大きさや身体ダイナミクスが異なるため生成された目標軌道をそのまま用いることはできない。これを初期目標軌道としてロボットはロボット自身に適した目標軌道を行動学習を通じて生成できるようになる。本研究では、倒立二輪型移動ロボットに人間の蹴球動作を観測させ、ロボットに適する目標軌道を生成する行動学習のシミュレーションを行い本手法の有効性を示す。

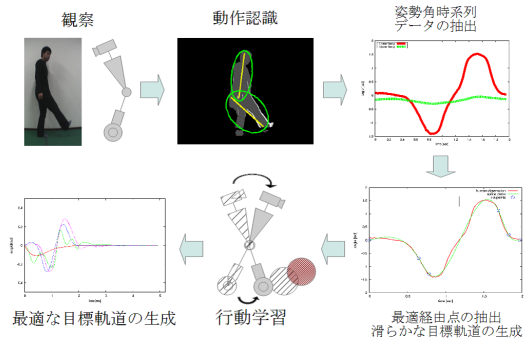


図1: 模倣学習フローチャート

2. 倒立二輪型移動ロボット

2.1 倒立二輪型移動ロボットの概要

図2(a)、(b)は本研究で用いる倒立二輪型移動ロボットのモデル、外観である。ロボットは左右平行二輪が接続する下半身と上半身の2リンクによって構成され、左右平行二輪と下半身、下半身と上半身の接点に1自由度ずつ持つ。 θ^l は下半身の姿勢角度、 θ^u は上半身の姿勢角度、 $\dot{\varphi}$ は車輪の回転加速度を表す。

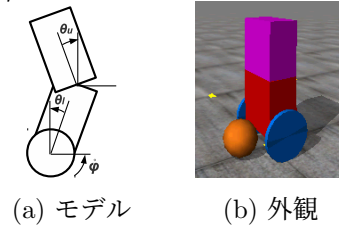


図2: 倒立二輪型移動ロボット

2.2 トルク制御式

倒立二輪型移動ロボットは、次の2つのトルク制御式で制御している。 u_1 は下半身と車輪のジョイント、 u_2 は上半身と下半身のジョイントにかかるトルクを表す。ロボットは下半身と上半身の目標姿勢角度 θ_d^l 、 θ_d^u を与えることによって前後に移動し動作を行うことができる。

$$u_1 = -k_1^l(\theta^l - \theta_d^l) - k_2^l\dot{\theta}^l - k_3^l\dot{\varphi} - k_4^l\sum(\dot{\varphi} - \dot{\varphi}_d) \quad (1)$$

$$u_2 = -k_1^u(\theta^u - \theta_d^u) - k_2^u\dot{\theta}^u - k_3^u\sum(\theta^u - \theta_d^u) \quad (2)$$

2.3 蹴球動作

本研究で倒立二輪型移動ロボットは、人間の蹴球動作を観察しその動作データを基に生成した目標軌道によって蹴球動作を実現する。ロボットがロボットの上下半身を前後に傾けることでロボットの下半身にボールを当てボールを転がす。上半身、下半身の傾きの動作は式 (1)、式 (2) に上半身、下半身の目標姿勢角度 θ_d^u 、 θ_d^l を入力することで実現する。

2.4 蹴球動作の評価

蹴球動作の評価は次のように設定し、評価関数を式 (3) に示す。この時、 l_r 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 はそれぞれ本体の積算移動距離、ボールの初速度の重み係数である。ロボットの積算移動距離がなるべく小さく、ボールの速度が速ければ評価が高くなる。また、蹴球動作後に倒立の安定性を評価するため、蹴球動作後の3秒間も評価を続ける。この3秒間に目標姿勢角度 0 をステップ入力し安定して倒立するかどうかを調べる。式 (3) に従うと蹴球動作後の振動が少ないほど、ロボットの積算移動距離が短くなり評価が高くなる。

$$E = -w_1 l_r + w_2 v_b \quad (3)$$

3. 目標軌道生成

3.1 滑らかな運動軌道の生成

倒立二輪型移動ロボットの動作は必要以上に振動することなく遂行されなければならない。必要以上に振動すると動作終了後の安定した倒立までの時間が増加したり、転倒して身体の破損に繋がるなどの問題がある。これを回避するには、振動が発生しにくい滑らかな運動軌道を入力として用いるのが適切である。しかし、人間の動作認識から得た運動軌道は細かく振動した滑らかな運動軌道でないことが多い。そこで、人間の動作認識から得た動作データからいくつかの経由点データを抽出しその経由点データを滑らかな曲線で補間する手法を用いる。この手法にはスプライン補間法を用いる。スプライン補間法は区分的に多項式関数（スプライン関数）を作り、与えられる点を滑らかにつなぎ合わせる方法である [3]。本研究では数あるスプライン関数の中でも良く使われる 3 次の自然スプライン関数を用いて経由点データ間の補間値を求め、滑らかな運動軌道を生成する。

3.2 最適経由点の探索

本研究では人間の蹴球動作を観測対象としており、高橋らの提案した動作認識法 [4] を用いる。人間の動作認識によって得られた動作データを図 3 に示す。横軸は動作時刻、縦軸は上下半身の姿勢角度である。この動作データ（動作時刻と姿勢角度）からいくつかの経由点データを抽出し人間の蹴球動作をスプライン補間によって再現する。この時、人間の呈示者の運動をより良く模倣するには、元となる提示者の運動軌道とスプライン補間によって生成した滑らかな運動軌道との誤差がより少ない方が適切であると考えられる。そこで、図 4 に示す最適経由点探索アルゴリズムに従って最適な経由点を求める。この時、倒立二輪型ロボットの動作の始点と終点は安定した倒立であることが望ま

しいため、スプライン補間の両端（始点、終点）の姿勢角度は $0[\text{rad}]$ とし、その間の経由点の最適な組み合わせを求める。経由点の必要十分な個数は任意に求められないため、1 個から選択していき十分に近似できる個数を調べる。呈示者の運動軌道とスプライン補間によって生成された運動軌道の誤差評価は、各時刻における姿勢角度の誤差を計算する。誤差の計算は式 4 によって行う

$$E = |\theta_s^2 - \theta_h^2| \quad (4)$$

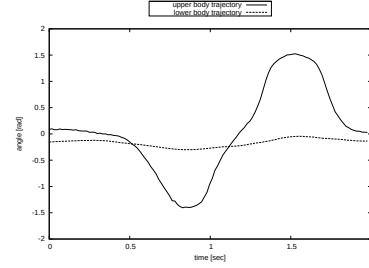


図 3: 人間の動作認識によって得られた動作データ [4]

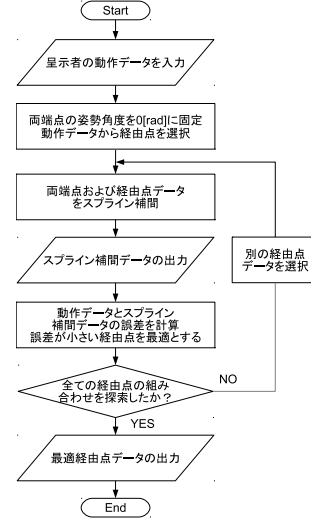


図 4: 最適経由点探索アルゴリズム

4. 最適経由点探索の結果

第 3.2 節の最適経由探索に基づき図 3 に示す人間の動作データ（運動軌道）から最適経由点を求める。人間の動作データは高橋ら [4] の実験データを用いる。人間の動作データを再現するために十分な経由点の個数は任意に求められない。従って、経由点の個数を 1 個から始め、1 個ずつ個数を増やして再現するのに最低限の経由点の個数を調べる。経由点の個数ごとの上下半身の最適経由点とその経由点をスプライン補間した結果を図 5(a)、(b) に示す。経由点の個数が増えるごとに行動学習のパラメータが増え、学習時間が増加するため経由点の個数は出来るだけ少ないのが望ましい。従って、図 5(a)、(b) より呈示者の運動軌道を再現する経由点の十分な個数は 4 個と考える。

5. 目標軌道学習

第 3.2 節で求めた最適経由点をスプライン補間法によって補間し呈示者の運動軌道に近似した滑らかな運



(a) 経由点 4 個下半身 (b) 経由点 5 個下半身

図 5: 最適経由点と両端点

動軌道を生成する。この滑らかな運動軌道を目標軌道として倒立二輪型移動ロボットに蹴球動作を行わせる。行動学習には方策勾配法を用いる。方策勾配法とは、より高い評価値が得られる方向に方策を繰り返し更新することで局所最適解を求める学習法である。本研究では Kohl and Stone によって提案された改良されたアルゴリズム [5] を用いる。行動学習のフローチャートを図 6 に示す。

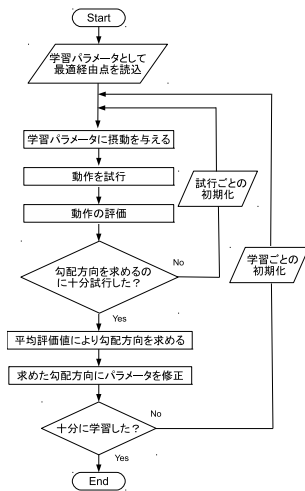


図 6: 目標軌道学習フローチャート

6. シミュレーション

シミュレーション環境は次のように設定し行動学習を行う。

6.1 学習パラメータ

この 4 個の経由点と探索の時に姿勢角度を 0 に固定した両端点（動作の始点と終点）を合わせたものを学習パラメータとする。学習パラメータは動作の始点と終点、その間の経由点（始点から見て経過時刻ごとに経由点 1、経由点 2、経由点 3、経由点 4 とする）で構成する。これらは、図 3 に示す人間の動作認識から得た動作データをスプライン補間によって再現する。ただし、人間の動作認識から得た姿勢角度の最大値は倒立二輪型移動ロボットが転倒せずに動作できる姿勢角度の値を大きく超えているため、入力する目標軌道（目標姿勢角度）のスケールは 5 分の 1 に縮小する。

6.2 学習環境の設定

試行時に各パラメータに与える小さな変動値は Table.1、学習パラメータを更新するステップサイズは Table.2 の通りに設定する。学習回数は 2000、1 回の学習に行われる蹴球動作の試行回数は 50 とする。

6.3 シミュレーション結果

行動学習の結果を以下に示す。図 8 は学習ごとの平均評価値の学習曲線を示す。平均評価値とは学習ごと

	上半身		下半身	
	時刻	姿勢角度	時刻	姿勢角度
動作の始点	0	0	0	0
経由点 1	0.002	0.001	0.002	0.001
経由点 2	0.002	0.001	0.002	0.001
経由点 3	0.002	0.001	0.002	0.001
経由点 4	0.002	0.001	0.002	0.001
動作の終点	0.002	0	0.002	0

表 1: 学習パラメータに与える変動値

	上半身		下半身	
	時刻	姿勢角度	時刻	姿勢角度
動作の始点	0	0	0	0
経由点 1	0.001	0.0005	0.001	0.0005
経由点 2	0.001	0.0005	0.001	0.0005
経由点 3	0.001	0.0005	0.001	0.0005
経由点 4	0.001	0.0005	0.001	0.0005
動作の終点	0.001	0	0.001	0

表 2: 学習パラメータを更新するステップサイズ

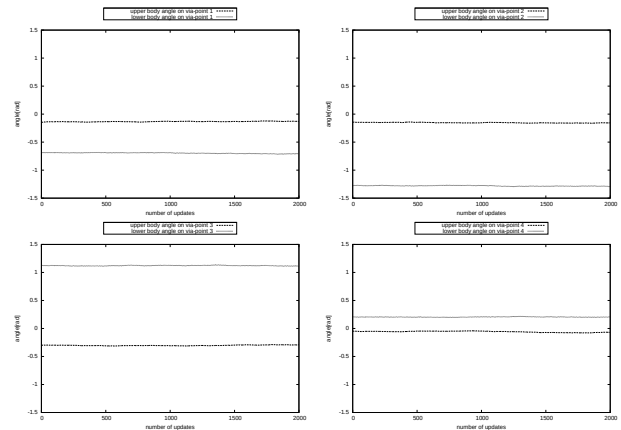


図 7: 上半身、下半身の経由点 1~4（姿勢角度）の推移に行う 50 回の試行の評価値の平均である。図 9 と図 10 は 2000 回学習まで行った試行のロボットの積算移動距離とボールの速度である。また、図 7 に学習ごとの一部の学習パラメータの推移を示す。図 11~図 14 に学習前、40 回、100 回、2000 回学習時に生成された上半身、下半身それぞれの目標軌道と実際のロボットの軌道を示す。

6.4 考察

図 9 と図 10 より、学習が進むごとにロボットの積算移動距離が短くなりつつボールの速度が速くなっていることから図 8 が示す通りに評価値が高くなっている。学習が進むごとの各経由点の姿勢角度や時刻の推移が微少にも関わらず評価は高くなっているため、微少な値でも学習が進むごとに蹴球動作を行うパラメータが最適化されていると考える。次に図 11~図 14 より目標軌道の推移をみると 1 秒目付近が山の頂点ととなっていた軌道が 0.5 秒付近で山の頂点になるという軌道に変化し、その山が徐々に大きくなっているのが分かる。一方で、学習が進むごとに蹴球動作後の下半身の振動が徐々に小さくなっているのが分かる。これは、上半身が早いうちに負の方向に傾くのが蹴球動作終了後の下半身の振動を小さくするのに関係していると考えら

れる。振動が小さいとロボットの移動距離が小さくなるため評価の向上に貢献していると考えられる。

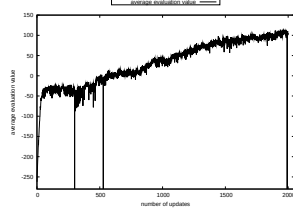


図 8: 平均評価値の学習曲線

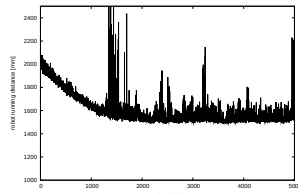


図 9: 試行ごとのロボットの積算移動距離の推移

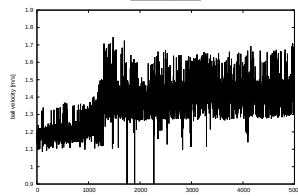


図 10: 試行ごとのボールの速度

7. 結 言

本研究では、人間の蹴球動作を倒立二輪型移動ロボットに動作認識させることでロボットがロボット自身に適する蹴球動作を獲得するという模倣学習をシミュレーションによって確認した。まず、ロボットが動作する際に入力する軌道はなるべく滑らかであるのが適切であることからロボットが動作する際の目標軌道として動作認識により得られた動作データをそのまま用いるのではなく、動作データから抽出したいくつかの経由点より再現した滑らかな軌道を動作の入力として用いる手法を提案した。この際、後の行動学習を行う際に学習コストが低くなるように最低限の経由点の個数を調べた。次に求めた最適経由点を用いて行動学習を行った。行動学習のシミュレーションの結果、学習が進むにつれてロボットは蹴球動作後の振動が小さくなるように目標軌道を変化させているのが分かった。これは動作後の倒立が安定しやすくなるように目標軌道を変化させていると考えられる。同時に蹴り出すボールの速度が増加していることから学習が進むにつれて動作パラメータの最適化が行われたと考えられる。今回の模倣学習はロボットに適する滑らかな目標軌道を蹴球動作後の倒立が安定するように学習が進行した。入力に用いる目標軌道のスケールはロボットの転倒を回避するために縮小したものとなっており、よりダイナミックで大きな動作の獲得は困難であると考えられる。これにより、ロボットの身体ダイナミクスの改良が今後の課題である。

参考文献

- [1] 細田祐司、中村亮介、柄川索、人間共生ロボットにおける倒立振り移動機構の走行制御方式、日本機械学会論文集 (C 編)、Vol. 76、No. 770、pp. 2678–2685 (2010-10-25).
- [2] 谷口忠大、岩橋直人、複数予測モデル遷移の N-gram 統計に基づく非分節運動系列からの模倣学習手法、知能と情報：日本知能情報ファジィ学会誌、Vol. 21、No. 6、pp. 1143–1154 (2009-12-15).
- [3] 川崎晴久、C & FORTRAN による数値解析の基礎、共立出版 (1993).
- [4] S. Takahashi, H. Nonoshita, Y. Takahashi and Y. Maeda, Inverted-pendulum Mobile Robot Motion Learning from Human Player Observation, SCIS & ISIS 2010 (2010).
- [5] N. Kohl and P. Stone, Policy gradient reinforcement learning for fast quadrupedal locomotion, in the *IEEE International Conference on Robotics and Automation*, pp. 2619–2624 (2004).

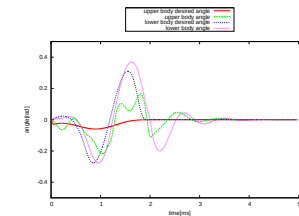


図 11: 学習前の目標軌道と実際の軌道

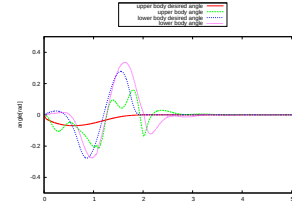


図 12: 40 回学習時の目標軌道と実際の軌道

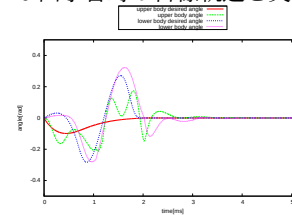


図 13: 100 回学習時の目標軌道と実際の軌道

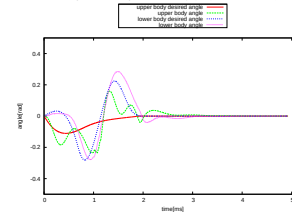


図 14: 2000 回学習時の目標軌道と実際の軌道