

A2-2 ヒューマノイドロボットのための対話型強化学習システムの開発

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

臼井 和廉 (指導教員：高橋泰岳, 前田 陽一郎)

1. 緒言

近年、人間との共生環境下において活動する自律ロボットの研究・開発が進められている。人間とロボットが共生するためには、ロボットが自ら環境に適応していくことが求められるが、人間の日常環境においてロボットに必要とされる全ての動作を事前に設計することは困難である。そこで、強化学習を用い、ロボットに試行錯誤を通して学習させる研究が行われている。試行錯誤による学習コストは大きく、特に実機ロボットに適用することは難しい。そのため、良い学習初期値を与えることが重要であり、人間や他のロボットの教示動作を学習初期値とする模倣学習の研究が多く行われている。また、同じ行動を学習する際にも動物らしい動きや人間の感じ方を評価に入れた学習の研究も行われている。

人間のスポーツや演技練習などの場合では、指導者は練習者の体格や理解に合わせて指導方法を変えていくことが一般的である。高橋らの研究 [1] では、車輪型のサッカーロボットの行動学習に他者行為の認識をフィードバックし、循環的に学習させることで行動学習が加速することを示している。そこで、本研究では、ヒューマノイドロボットのための対話型模倣学習手法を提案し、その実装を行う。

2. 対話型模倣学習

学習後のロボットの動作を評価し、ロボットの動作がより良くなるよう教示動作の修正を行い、さらにロボットに学習させていく (図 1)。まず、人間が教示動作を呈示する。そして、ロボットを学習中の動作パラメータで動作させ、人間とロボットのリンク姿勢を比較し、ロボットのリンク姿勢が人間のリンク姿勢に近くなるよう動作パラメータを方策勾配法により学習させる。学習後のロボットの動作から教示動作を修正し、再び教示動作の呈示を行う。この手順を繰り返すことによって対話型模倣学習を実現する。

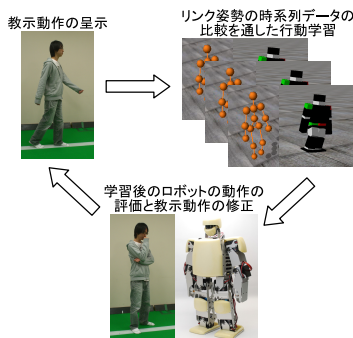


図 1: 模倣学習の流れ

3. 動作認識

人間の教示動作を認識するために、モーションキャプチャを行い教示動作データを取得する。ここでは、Microsoft 社の Kinect センサ (以下、Kinect) を用いてモーションキャプチャを行い、マーカや他のセンサ類を装着せずに人間の骨格を検出し、関節座標を記録する。

Kinect を用いて人間の 16 点の関節座標を取得し、16 点の関節座標をつなぎ 11 個の関節リンクを作成する (図 2)。取得した関節座標は、Kinect の位置を原点としたカメラ座標系 Σ_c での値となっているため、ロボット座標系 Σ_r への座標変換を行う。ロボット座標系は、腰座標を原点とし、左腰座標から右腰座標へのベクトル方向を X_r 軸、地面に鉛直上向き方向を Z_r 軸とし、 Z_r 軸と X_r 軸の外積方向を Y_r 軸とする (図 3)。

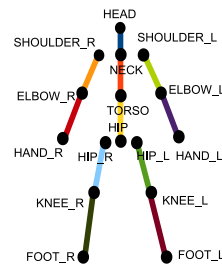


図 2: Kinect で取得する関節とリンク

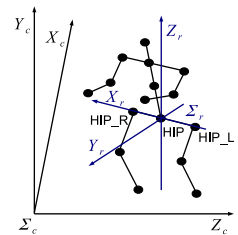


図 3: カメラ座標系とロボット座標系

取得した人間の教示動作とロボットの動作を比較し、動作認識を行うために各リンク i の姿勢 φ_i を用いる。ここで、リンク姿勢とは関節リンクの単位ベクトルとする。人間のリンク姿勢 ${}^h\varphi_i$ とロボットのリンク姿勢 ${}^r\varphi_i$ の内積の和をリンク姿勢の類似度 S として算出し、比較する (式 (1))。 N は関節リンク数、 T は関節角軌道の時系列データ数である。

$$S = \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N {}^h\varphi_i(t) \cdot {}^r\varphi_i(t) \quad (1)$$

4. 経路点を用いた関節角軌道表現

関節角軌道の時系列データは、データ数が多くそのままロボットの学習パラメータとするには適さない。そこで、木村らの研究 [2] では、運動軌道の時系列データからいくつかの経路点を抽出し、その経路点を学習パラメータとすることで学習パラメータのデータ数を削減している。本実験で用いるヴイストン株式会社のヒューマノイドロボット、Robovie-PC の動作は、三次曲線補間を用いて関節角軌道を生成していることから、三次曲線補間を用いる。

5. 行動学習

ロボットが自身に適した動作を獲得するために行動学習を行う。本研究では、行動学習の手法として強化学習の一種である方策勾配法を用いる。方策勾配法として様々なアルゴリズムが提案されているが、Kohl and Stone によって定式化されたアルゴリズム [3] を利用する。

本実験では、歩行動作の模倣学習を行う。関節角軌道の経由点を学習パラメータとし、評価値 E_L を評価式 (2) より求め、学習を行う。ここで、 D はロボットの歩行距離、 w_S 、 w_D はそれぞれリンク姿勢の類似度 S と歩行距離 D に対する重み係数であり、 P_L は転倒時のペナルティである。評価式 (2) では、ロボットと人間のリンク姿勢の類似度が高く、歩行距離が大きいほど評価は高くなり、ロボットが転倒してしまうと評価は低くなる。

$$E_L = w_S S + w_D D - P_L \quad (2)$$

6. 実験

人間の動作を Kinect を用いてフレームレート 30[fps] で記録し、物理計算エンジンライブラリである ODE (Open Dynamics Engine) で開発したシミュレータを用いて、ロボットの歩行動作の行動学習を行った。

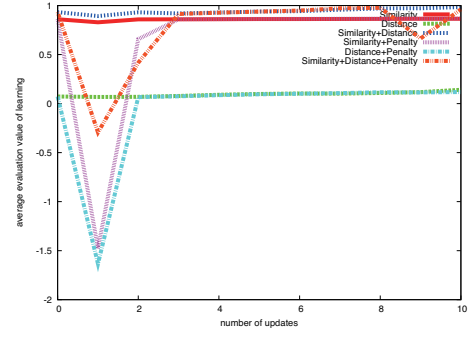
試行回数を 100 回、学習回数を 10 回、学習パラメータ更新時のステップサイズを 0.5° 、動作評価における転倒時のペナルティ $P_M = 10$ として表 1 に示す 6 通りの条件でシミュレーションを行った。その結果を図 4、図 5 に示す。どの条件でも、1 回目の学習では転倒回数が多く評価値は低くなっているが、最終的に学習前よりも評価値は高くなった。ロボットの歩行距離のみを評価して学習した条件 2 は、転倒回数が多い。それに比べ人間の教示動作を学習に取り入れた場合では、転倒回数が少なく安定した歩行動作を学習できたと考えられる。

表 1: 実験条件

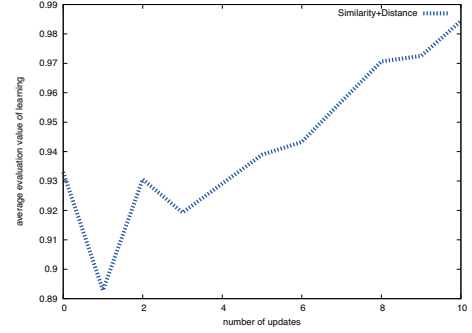
	w_S	w_D	P_L
条件 1	1.0	0.0	0.0
条件 2	0.0	1.0	0.0
条件 3	1.0	1.0	0.0
条件 4	1.0	0.0	10.0
条件 5	0.0	1.0	10.0
条件 6	1.0	1.0	10.0

7. 結言

本研究では、対話型模倣学習を提案した。また、リンク姿勢を比較し、ロボットのリンク姿勢が人間のリンク姿勢に近づくよう学習させることでロボットに人間の動作を模倣させた。今回は、1 回の教示動作に対して模倣学習を行いその有効性を示した。今後、今回実装したシステムの効率化を進め、対話型模倣学習の実験を行い、ロボットの学習だけでなく人間の教示動作の変化についても検証していく。



(a) 条件 1~6 の評価値 E_L の平均値



(b) 条件 3 の評価値 E_L の平均値

図 4: 評価値 E_L の平均値

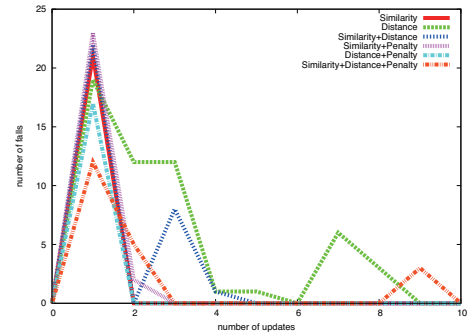


図 5: 転倒回数

参考文献

- [1] Yasutake Takahashi, Yoshihiro Tamura, Minoru Asada, Mario Negrello, "Emulation and behavior understanding through shared values," *Robotics and Autonomous Systems*, Vol.58, pp.855-865 (2010)
- [2] 木村 竜也, 高橋 泰岳, 前田 陽一郎, 中村 恭之, "人間動作の経由点表現を用いた倒立二輪型移動ロボットによる模倣学習," 第 28 回ファジィシステムシンポジウム, WE2-2, Vol.CD-ROM, pp.458-463 (2012)
- [3] Nate Kohl, Peter Stone, "Policy Gradient Reinforcement Learning for Fast Quadrupedal Locomotion," *Proceedings of the IEEE International Conference on Robotics and Automation*, pp.2619-2624 (2004)