ジェスチャによるロボット動作訓練に関する研究

北見工業大学　○小川弘　鈴木育男　岩館健司　渡辺美知子

要旨

本研究では，専門知識のないユーザーがロボットに対して，容易に学習を行わせるための手法として，人間のジェスチャを利用した学習手法を提案する．ジェスチャの認識にはKinectを利用し，学習方法には強化学習の一つであるQ学習を用いる．実験では複数の状態（ジェスチャ）に対して，それぞれ対応した動作を学習させ，そのQ値，行動選択率の遷移により，手法の有効性を示す．

1. はじめに

近年，ますますロボットは我々の生活圏の中に浸透しつつあり，専門知識のないユーザーでもロボットに対して学習が行えるように簡便な動作設計（訓練）の手法が求められるようになってきている．本研究では人間のジェスチャを利用し[1]，ロボットの動作訓手法の獲得を目的とする動作訓練の方法として強化学習法を用いる．

2. 提案手法

提案システムのフローチャートをFig. 1に示す．

本研究では，学習手法にQ学習を用いる[2]．Q学習とは強化学習の一つで実行する各動作に対して，その有効性を表すQ値を持たせ，その値を実行毎に更新していき，状態と行動の結びつきを強化する学習法である．

ロボットには二輪走行をするLEGO Mindstorms EV3を使用した．ジェスチャを認識するためのKinect・PCとEV3との通信にはBluetoothを用いる．

本研究では状態を人間のジェスチャ，行動をロボットの動作とし，前進，後退，右回り，左回りの四種類の運動を評価する．以下に提案システムの詳細について説明する．

2.1. 初期化

各状態における四つの動作に対して，それぞれのQ値を全て0.0に初期化する．

2.2. ジェスチャ入力および動作決定

人間のジェスチャの判別にはKinectから得られる骨格情報を利用する．特定のジェスチャを一定時間続けることによりジェスチャから状態を認識する．

ロボットの動作を決める手法としてソフトマックス手法を用いる．式(1)のソフトマックス手法はQ値の収束を早め，より大きな値を持つ動作が選ばれるように各Q値のエクスポネンシャルの合計範囲で取る乱数により行動を決定する手法である．

$$\pi(s_i, a_i) = \exp(Q(s_i, a_i)/T) \Big/ \sum_{j \in A} \exp(Q(s_i, j)/T) \qquad (1)$$

ここで，Tは正の定数，$S=\{s_1, s_2, s_3, \cdots, s_M\}$は状態集合，$M$は状態数，$A=\{a_1, a_2, a_3, \cdots, a_N\}$は行動集合，$N$は行動数，$j$は状態$s_i$で可能な行動集合，$Q$は各$Q$値である．

ロボットはソフトマックス手法により生成した乱数に対応した範囲にあるQ値を持つ動作を行う．

2.3. 行動の評価

評価には人間のジェスチャを用いる．ジェスチャはKinectによって認識し，両手を挙げた状態を高評価，それ以外を低評価として行動の報酬を決める．

2.4. 学習・更新

動作の評価を基にQ学習を行う．Q値の更新は式(2)で行う．ここで，αは学習率であり，Rは報酬である．

$$Q(s,a) = (1.0 - \alpha)Q(s,a) + \alpha R \qquad (2)$$

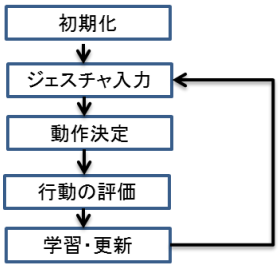


Fig. 1 システムフローチャート

3. 動作訓練実験

本研究では行動数を共通してN=4とし，状態が一つがである単一状態実験と複数である複数状態実験の2つを行った．また，複数状態実験では状態数が増加することによる学習時間も同時に増加するため，対策として評価によるQ値の更新に変更を行った．単一状態実験においては高評価が与えられた場合にのみ報酬を与え，Q値に変化を起こすが，複数状態実験では取った動作が目的のものでなく，高評価を得られなかった場合，低評価とし負の報酬を与え，その動作が選ばれにくくする．

各実験における報酬と学習率Table. 1にまとめる．

Table. 1 各実験の報酬と学習率

	報酬	学習率
単一状態実験	2	0.1
複数状態事件	高評価:2 低評価:-1	0.2

3.1. 単一状態実験

はじめに，Q学習の有効性を確認するため，単一の状態において，Q値の遷移を測定する．後退動作の獲得を目的とし状態数M=1，後退動作のみを評価し，5回連続して後退動作を行うことによって学習を完了したものとする．単一状態のため，ジェスチャ入力は省く．

3.2. 実験結果

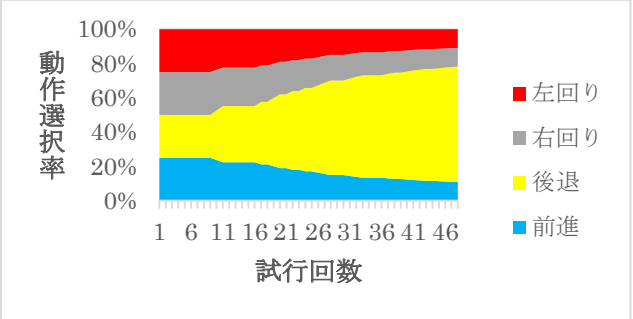


Fig. 2 Q学習時の行動選択率の遷移

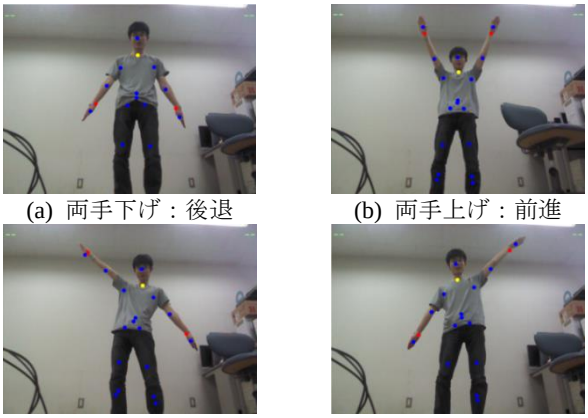
約50回程の試行を重ねて学習を完了した．Fig. 2に学習完了までの各動作の選択率の遷移状況を示す．グラフから，学習前は全ての動作の選択率は一律だが，最終的に後退のQ値の割合が約70%に達し，十分な選択率となっていること

がわかる．

3.3. 複数状態実験

前実験で省いたジェスチャ入力を実装する．ジェスチャから状態を認識し，複数の状態における Q 学習を実現する．

状態数 M=4 とする．ジェスチャの種類として両手を上げた状態，両手を下げた状態，右目のみを上げた状態，左手のみを上げた状態を設定した．各状態に対しそれぞれ，前進，後退，右回り，左回り，の内一つを目的動作とし学習させる．使用するロボット，及び学習を完了する条件は前実験と同様とした．Fig. 3 に各ジェスチャとその状態において学習を行う目的の動作を示す．



(a) 両手下げ：後退　(b) 両手上げ：前進

(c) 右手上げ：右回り　(d) 左手上げ：左回り

Fig. 3 各ジェスチャ：目的動作

3.4. 実験結果

Fig. 4-Fig. 7 に各状態における学習完了までの Q 値の遷移を示す．グラフから，学習完了までの試行回数や遷移の仕方に違いはあるが，いずれの状態にも最終的には目的動作の選択率は 60%から 70%程に達した．また，改善した Q 値の更新によって目的以外の動作が選択された際にも選択率に緩やかな変化が発生し，前実験と比べて学習効率が向上している．

4. まとめと今後の課題

2 つの実験の結果から以下の結論が得られた．

(1)Q 学習によりジェスチャを利用した複数の状態における各動作の学習に成功した．

(2)Q 値の更新の改善により学習効率が向上した．

今後の課題として，現在ではあらかじめ設定したジェスチャしか認識できないこと，認識できるジェスチャ，状態数の増加に伴ってさらに学習時間を要することが挙げられる．学習時間の問題に関しては Learning from Easy Missions [3]等の手法があるが，本研究に適用できるか今後検討していきたい．Q 値の更新式も過学習とならない範囲で，学習率や報酬を変えていき，より効率の良いものへと改善していく．また，現在は単一の動作のみだが，連続した動作の学習，評価を行えるようにしていく．

また，二輪走行のロボットでは行える動作が単調であるため，使用するロボットもより複雑な動きができるものへと変えていく．ジェスチャによる評価もリアルタイムで行えるようにしたい．

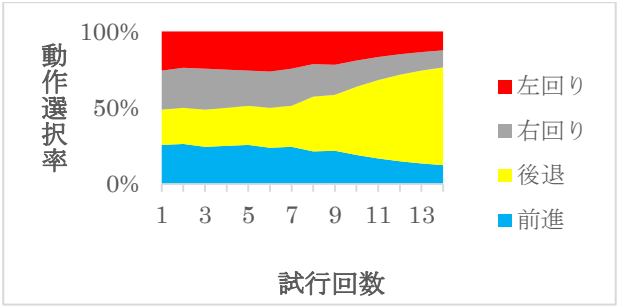


Fig. 4 行動選択率の遷移：両手下げ

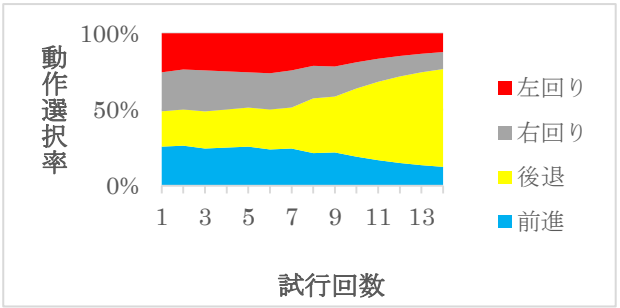


Fig. 5 行動選択率の遷移：両手上げ

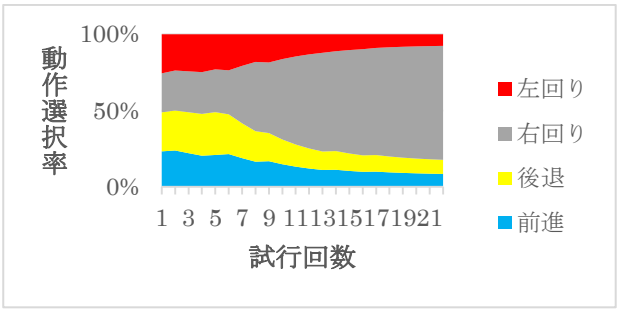


Fig. 6 行動選択率の遷移：右手上げ

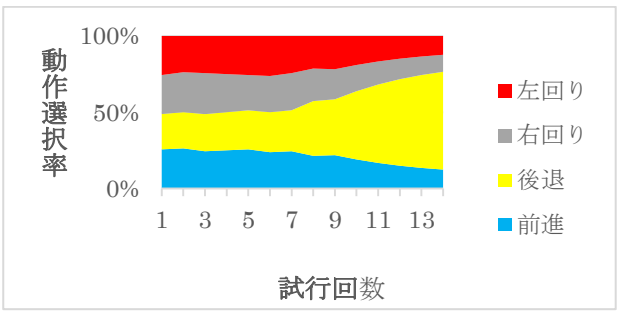


Fig. 7 行動選択率の遷移：左手上げ

参考文献

[1]吉田成朗，鳴海拓志，橋本直，谷川智洋，稲見昌彦，五十嵐健夫，廣瀬通孝，“ジェスチャ操作型飛行ロボットによる身体性の拡張”，情報処理学会 インタラクション, 2012.

[2]柴田克成 “強化学習とロボット知能 - あめとむちで知能は作れるか？ - ”，第 16 回人工知能学会全国大会論文集，パネルディスカッション「強化学習とその諸相」パネリスト原稿, 2A1-05, 2002.

[3]野田彰一，浅田稔，俵積田健，細田耕，”強化学習によるロボットの行動獲得の効率化に関する考察—簡単タスクからの学習 LEM—”，日本ロボット学会ロボットシンポジウム予稿集 , 4 巻, pp67-72, 1994.