

強化学習による人間位置に基づいたロボットの挙動選択

田崎豪

駒谷和範

尾形哲也

奥乃博

京都大学大学院情報学研究科知能情報学専攻

1. はじめに

近年多くのロボットが開発され、ロボットが複数の人とインタラクションを行う機会が増えてきた。我々は、ロボットが複数の人とインタラクションを行った場合でも、それぞれの人に分かりやすく、親しみやすい印象を与えることを目的とする。複数の人とインタラクションをとる場合、ロボットは人の位置を把握し、インタラクションを行うべき人を選択し、選択した人に適切な動作を行う必要がある。我々は、人を把握し、選択する手法として、親密度空間マップを設計した [1]。本稿では、選択した人に適切な動作を行う手法について述べる。

適切な動作をとらせるため、光永ら [2] は、ロボットに一定の動作量と動作速度を学習させた。しかし、学習した「一定の」動作が適切とは限らない。筆脇ら [3] は、人とロボットの位置関係に応じて、人が感じる快、不快感が増えることを実証した。したがって、光永らのモデルにおいてロボットの動作量、動作速度は一定ではなく、位置関係に応じて変化する必要がある。例えば、人が遠くにいる場合、動作量の大きなロボット動作が好まれるかもしれないが、近い場合は好まれないかもしれない。

我々は人位置を考慮した強化学習を設計し、位置に応じて動作量、動作速度を変更する手法を開発する。人位置を考慮することで、ロボット動作の危険性を減らし、ロボット動作に対する印象の向上が期待できる。

2. 動作情報学習空間の設計

2.1 親密度空間マップ

現在のセンサでは実環境下で正確な人定位が困難であり、人も cm 単位で相手の位置を把握しているわけではない。ロボットもインタラクションを行う空間を分割し、分割された空間程度の範囲で人を把握できればよい。我々はどの程度分割するかを決定するため、近接学 [4] とロボットモジュールの有効距離と人の顔の大きさ等を考慮した。近接学におけるコミュニケーション距離と実験的に得られたロボットモジュールの有効距離の関係は表 1 のようになる。近接学では、相手との関係に応じてコミュニケーションをとる距離は変わるとされ、親密距離は 50cm 以内、個人距離は 50-120cm、社会距離は 120-360cm となっている。

表 1: コミュニケーション距離とモジュール

親密距離	個人距離	社会距離
触認識		
音声認識	音声認識	
顔定位	顔定位	顔定位
音源定位	音源定位	音源定位

表 1 を考慮して設計されたインタラクションを行うための空間が親密度空間マップである (図 1)。親密度空間マップの分割空間上で人を把握すれば、実環境下で人を把握、選択し、印象のよい動作が行えることが確認されている [1]。

Robot Motion Selection Based on the Position of People by Using Reinforcement Learning: Tsuyoshi Tasaki, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

2.2 強化学習の設計

本研究では適切な動作量、動作速度を学習するため Q-Learning を用いる。Q-Learning はインタラクション中に比較的速く収束し、設計が容易であるという利点がある。

親密度空間マップ上の部分空間を Q-Learning で定義する状態とみなし、動作量と動作速度について学習を行い、その結果を部分空間にマッピングする。近接学、センサ能力を考慮して設計された状態を使用することで、効果的な学習と、位置に応じた動作情報変更が可能になる。各状態で選択できる行動は、動作量と動作速度の組み合わせ 4 種類、つまり (動作量、動作速度)=(小、遅)、(小、速)、(大、遅)、(大、速) とする。ここでは、親密度空間マップ上で選択された人の位置が移動したとき、状態が遷移したとみなす。

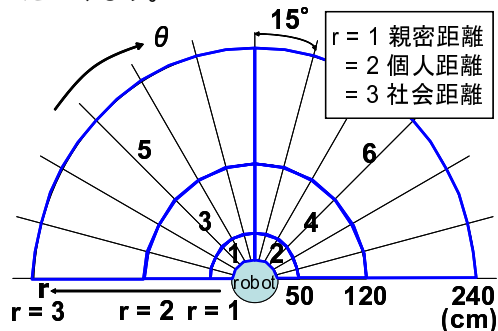


図 1: 親密度空間マップと学習に用いた状態 (空間)

状態を図 1 で示した全てのセルに設定すると学習に時間がかかり、その多様性から被験者がロボットの行動を解釈しづらくなる。したがって、動作情報の学習には図 1 の太線で示す 6 状態 (空間) を用いることにした。ここでは動作変更実験の第一歩として対象人数を二人と仮定し、状態数を決定した。

本研究ではロボットが人にとって親しみやすい動作を行うことが目的である。ロボットは、人を親密距離にひきつけ、快い刺激を与えられることが望まれる。したがって報酬は、人がロボットに近付いたとき、人が親密距離内に存在するとき、握手されたという快刺激を受けたときに与えられる。罰は、人がロボットから遠ざかったとき、人が親密距離外に存在するとき、蹴られたという不快な刺激を受けたときに与えられる。

学習率を lr 、割引率を dr 、得られた報酬を re とすると、空間 i ($i=1, 2, 3, 4, 5, 6$)、行動 u ($u=1, 2, 3, 4$) の、時刻 t における Q 値は式 (1) で更新される。ここでは報酬を重視し、 $lr=0.5$ 、 $dr=0.9$ とした。

$$Q(i, u)_{t+1} = (1 - lr) * Q(i, u)_t + lr * dr * re \quad (1)$$

ロボットが各空間の動作量、動作速度を学習する過程を以下に示す。

1. マップで選択された人がいる空間 (状態) を検出する。
2. ϵ -greedy 法により、行動を 1 つ選択し、選択された行動をもとに動作量、動作速度を設定する。
3. 行動した結果得られた報酬によって、選択された行動の Q 値を式 (1) に基づいて更新する。

3. 人間位置に基づくインタラクションの設計

3.1 ヒューマノイドロボット RobovieR2

本研究のテストベッドとして、図2の中央に示すヒューマノイドロボット RobovieR2を使用した。RobovieR2の入力センサは、肩と腕、手先、足元に装着されたタッチセンサ、頭部にある1対ずつのマイクと全方位カメラである。出力として、スピーカと3自由度の首のモーター、4自由度ずつの左右腕のモーターを用いる。RobovieR2には入力モジュールとして触認識、顔定位、音源定位が実装されている。触認識は、触られた場所から「握手された」「蹴られた」を識別し、快不快を判断する。また出力モジュールとして、接触応答、抱擁動作、手振動作、親密人間選択動作、顔方向追跡動作、音源方向追跡動作が実装されている。



図2: RobovieR2 との実験風景

3.2 インタラクション中のロボット動作

- マップに基づくロボットの具体的な動作を以下に示す。
1. 触認識、顔定位、音源定位が検出した方向へ、それぞれ接触応答、顔方向追跡、音源方向追跡を用いて動作する。各動作は人を検出した空間にマッピングされた動作量、動作速度を使用する。
 2. 親密度空間マップ参照後、人定位結果と快不快に応じてマップの更新と動作情報の学習を行う。
 3. 親密距離内でマップが示す親密度 [1] が閾値より高い空間を検出した場合は抱擁動作を行う。親密度が閾値よりも高い空間が個人距離、社会距離にあれば手振動作を行う。該当する空間が無ければ親密人間選択でマップ上最も親密度が高い空間を凝視する。

4. 人間位置に基づいた挙動選択の印象評価

4.1 評価実験の目的と内容

人の位置に応じて動作を選択することが、好印象を与えているかをアンケート調査した。アンケート用紙には、[1]で因子分析により得られた印象評価用形容詞(表2)が並べてあり、どの程度その形容詞にあてはまるかを7段階評価してもらう。表2では、第一因子を構成する形容詞が太字で、第二因子を構成する形容詞が下線で示されている。被験者は21-27歳の男女2人ずつ計8名であり、それぞれのグループに RobovieR2 とインタラクションをとってもらった。

インタラクションは2回行い、RobovieR2はそれぞれの試行で異なる行動方式を用いる。一方は、節3.2で示した行動方式である。もう一方は、ベースラインとして位置を考慮しない行動方式を実装した。近くで学習したパラメータを遠くでも使用し、距離に関係した報酬、罰は与えられない方式である。この方式では距離を考慮しないため、距離に関連する抱擁動作と手振動作を行わない。

試行の前に、ロボットが触認識、顔定位、音源定位モジュールを有し、動作量と動作速度を変更させることの

みを伝えておく。人位置を考慮することや具体的な動作原理は説明しない。各試行とも10分インタラクションを行う。10分のうち7分はタッチセンサを中心に自由なインタラクションをとってもらい、残りの3分は親密距離の外側から中心にインタラクションをとってもらう。近距離の場合も遠距離の場合も、RobovieR2の動作量と動作速度が自分にとって特に適切な場合には握手をしてもらい、特に不適切な場合には蹴ってもらった。

表2: アンケート用形容詞(7段階評価)

やさしい	感じのよい	親しみやすい	安全な
暖かい	かわいらしい	うちどけた	わかりやすい
近づきやすい	明るい	思いやりのある	人間的な
充実した	面白い	愉快的な	好きな
興味深い	良い	複雑な	速い
すばやい	はげしい	積極的な	強気な
派手な	陽気な	敏感な	賢い

4.2 実験結果と考察

各行動方式における、アンケートから得られた形容詞の評点平均を表3に示す。t検定の結果、有意水準0.01で両因子を構成する形容詞の評点平均が、人位置を考慮した場合に高くなることが確認された。

表3: 形容詞評点平均の比較

	第1因子		第二因子	
	平均	標準偏差	平均	標準偏差
人位置考慮	4.72	0.79	4.97	1.24
人位置無視	3.97	1.02	4.13	1.11

得られた動作量、動作速度は被験者グループ、距離ごとに異なっていた。したがって、人ごとに、距離に基づく動作情報学習は必要であると考えられる。ロボットが人位置を考慮して適切な動作をしたため、人にとって受け入れやすい動作になったものと考えられる。

グループごとに得られた動作情報は上記のように異なっていたが、同時に実験した被験者グループの中では距離ごとにほぼ同様の学習結果が得られていた。これは人が集団でインタラクションをとる場合の引き込み現象が関係していると考えられる。したがって今後は、人が2人いる場合でも、学習する空間を左右方向に分けず、3状態で学習を行うことでより速く学習が行えると考えられる。また、互いの干渉が少なかった一組だけ、異なる動作情報が得られていた。3人以上のインタラクションでは干渉しない人が増えると考えられる。したがって、引き込みあっているグループと、干渉しないグループのそれぞれに対する挙動選択法を考案する必要がある。

5. おわりに

本稿は、インタラクション時のロボット動作選択法として、インタラクション中の人位置に着目した、強化学習による動作変更法を報告した。本手法により、ロボット動作に対する人の印象向上を確認した。今後は、対話も含めた動作内容についての挙動選択も行いたい。謝辞 本研究の一部は、日本学術振興会科学研究費補助金、21世紀COEプログラムの支援を受けた。

参考文献

- [1] Tsuyoshi Tasaki, et al.: "Spatially Mapping of Friendliness for Human-Robot Interaction", IROS, pp.521-526, 2005.
- [2] Noriaki Mitsunaga, et al.: "Robot Behavior Adaptation for Human-Robot Interaction based on Policy Gradient Reinforcement Learning", IROS, pp. 1594-1601, 2005.
- [3] 筆脇晃朗ら: "パーソナルスペースを考慮した人型ロボットの間合い制御に関する研究", ROBOMECH, pp.1P1-S-008.1~2, 2005.
- [4] Hall, E. T.: "Hidden Dimension. Doubleday Publishing", ISBN 0385084765, 1996.