

複数タスクに対するロボットの行動学習 -タスクの重要度に基づく行動選択手法の提案-

室蘭工業大学 情報電子工学系専攻 認知ロボティクス研究室 三浦 丈典

1. はじめに

近年、ロボットの性能は大きく向上し、1つのタスクを実行するだけでなく、複数のタスクを扱うことのできる汎用的な自律ロボットが増えている。複数のタスクを実行可能なロボットには、様々なタスクが同時に実装されることが想定される。従って、ロボットは複数のタスクの優先関係を考慮し、それぞれのタスク遂行を考える必要がある。

2. 従来研究

複数のタスクが実装される自律ロボットにおいて、強化学習を適用することで、複数タスクを遂行する研究が行われている。例としては、重み付き報酬関数を用いた研究[1]や多目的最適化問題を適用した研究[2]等がある。これらの研究では、各タスクの重要性からタスク間の優先関係を設定し、それを基に各タスクに重み付けを行なうことで行動選択を行っている。

3. 従来研究の問題点

従来研究では、タスクの優先関係を重みとして表現し、予め設定しておくことで複数タスク下での行動選択を行っていた。しかし、実際の環境においてタスク間の優先関係は変化する可能性がある。ロボットの状態や、周囲の環境によって各タスクの重要性は変化する。個々のタスクの重要性が変化すれば、タスク間の優先関係も変わってしまう。

従来研究では、ロボットが行なうタスクやその優先関係は、学習中に変化しないことを前提としていた。従って、学習中にタスクの優先関係が変化した場合には、その度に最適化を行なう必要がある。

4. 本研究の目的

本研究では、複数のタスクが与えられるロボットにおいて、タスク間の優先関係が変化した場合でもその変化に対応可能な行動学習・行動選択手法を提案する。

5. アプローチ

本研究では、ロボットの状態や周囲の環境による各タスクの重要性の変化に対応するために、ロボットが認識する状態に応じて各タスクの重要性を算出する。また、各タスクについての行動学習については、タスク毎に別々の学習空間で行なう。その後、行動学習により蓄積した行動価値と、各

タスクの重要性に基づき行動選択を行なう。行動学習と各タスクの重要性による行動選択を分けることで、タスク間の優先関係が変化した場合にも、再学習を行わず行動選択を行なうことが可能となる。

6. 複数タスク下での行動選択手法

本研究では、状態に応じて変化する各タスクの重要性を表す指標を重要度と定義する。重要度は $-1 \sim 1$ の範囲の値で、この値が、 $+1$ に近づけば重要性が高いことを意味し、 0 に近づけば重要性が低いことを意味する。また、 -1 に近づくと重要性は無く、他のタスクを優先する。

ロボットが直面する状態における各タスクの重要度は(1)式によって算出する。この関数 $f(s)$ は、ロボットの状態を引数とする関数で、引数となる状態はタスク毎に設定する。

また、重要度 p_i は(2)式によって更新される。

$$f(S) = \frac{-2}{1+e^{(\delta \times S + \sigma)}} + 1 \quad (1)$$

δ, σ : は定数

$$p'_i \leftarrow p_i + \theta \{f(S) - p_i\} \quad (2)$$

提案手法では、各タスクに対して別々に行動学習を行うが、タスク間の行動価値の範囲を統一するために報酬関数の取りうる値は $-1 \sim 1$ に正規化する。次に、蓄積した行動価値と算出した重要度から行動選択を行なう。まず、各タスクの行動価値を軸とするタスク間行動価値空間を構成する。構成した空間上に、行動 a_i と算出した各タスクの重要度の交点 P' をプロットし、各行動と重要度の交点とのユークリッド距離 d_i が最も短い行動を選択する(図1)。

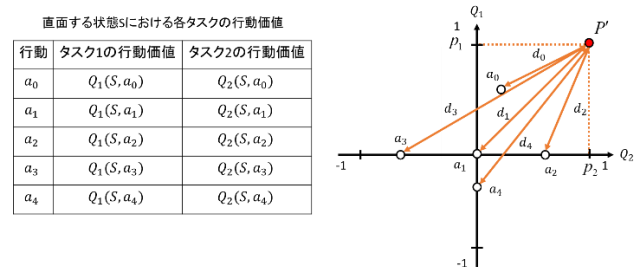


図1 行動価値と重要度を基にした行動選択

7. シミュレーション実験

提案手法を適用したロボットが、各タスクの重要度に応じた行動選択が可能であることを示した

めにシミュレーション実験を行う。実験では、2つの競合するタスクを与え、行動学習中に各タスクの重要度を変化させる。

実験環境は、図2に示す。ロボットは前後左右、静止の5つの行動が可能で、行動する度にエネルギーを消費する。

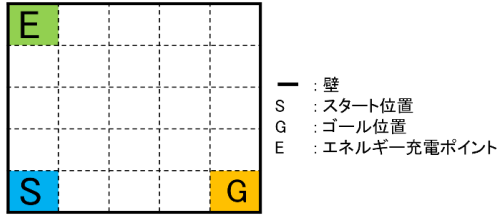


図2 実験環境

ロボットには、エネルギー獲得タスクとゴール到達タスクの2つを与える。エネルギー獲得タスクは、(3)式によって、1回の行動で獲得するエネルギー変位量を基に報酬を与える。ゴール到達タスクは、(4)式によって、ロボットがゴールへ到達した場合に報酬を与える。

$$r_1 = \Delta E \quad (3)$$

$$r_2 = \begin{cases} 1(\text{ゴール位置}) \\ 0(\text{その他}) \end{cases} \quad (4)$$

エネルギー獲得タスクの重要度は、エネルギー残量を基に(5)式によって設定する。また、ゴール到達タスクについては、人間のためのタスクであるので、人間がタスクの重要性に応じて-1~1の値 r_h を与える。この r_h を基に(6)式によって重要度を算出する。

$$f(E) = \frac{-2}{1+e^{(\delta \times E + \sigma)}} + 1 \quad (5)$$

δ, σ : 定数 E: エネルギー残量

$$f(r_h) = \frac{-2}{1+e^{(\delta' \times r_h + \sigma')}} + 1 \quad (6)$$

δ', σ' : 定数,

r_h : 人間がタスクの重要性に応じて与える値

r_h については、ロボットがゴールする度に与え、値については10000回行動する毎に、1, 0.3, -1の順に変化させる。

実験で用いた各種パラメータを表1に示す。

表1 実験パラメータ

行動回数	30000回	学習率 α	0.1
初期行動価値	0.0	割引率 γ	0.9
ランダム行動の確率	0.05	θ	0.5
ロボットの初期エネルギー残量	100	μ	0.5
最大エネルギー残量	100	δ	-0.13
最低エネルギー残量	0	σ	10
充電ポイントでのエネルギー変位	+3.0	δ'	5
移動時のエネルギー変位	-0.25	σ'	0
静止時のエネルギー変位	-0.05		

8. 実験結果と考察

ロボットの行動回数に対するエネルギー残量とゴール回数の推移を図3に示す。また、そのときの各タスクの重要度の推移を図4に示す。

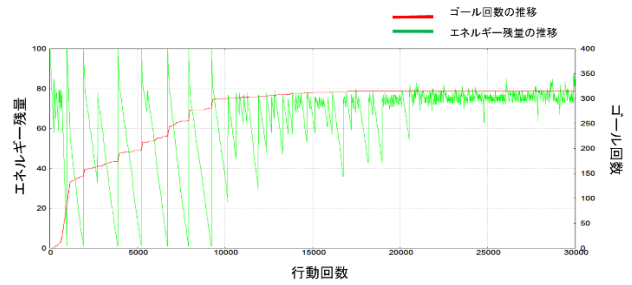


図3 行動回数に対するエネルギー残量とゴール回数の推移

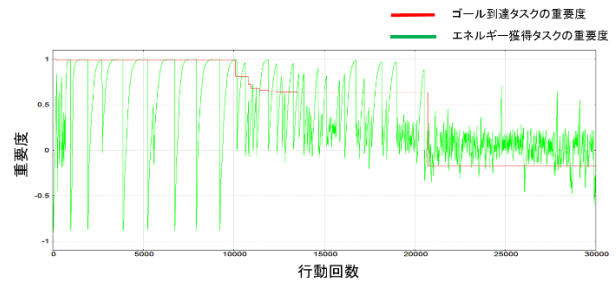


図4 行動回数に対する各タスクの重要度の推移

実験結果から、ゴール到達タスクの重要度が高く、エネルギー獲得タスクの重要度が低い場合には、エネルギー残量に関わらずゴール位置へ向かっている。従ってゴールへの到達を優先している。また、ゴール到達タスクの重要度が低く、エネルギー獲得タスクの重要度が高い場合には、ゴールへはあまり向かわずエネルギー残量80付近で維持されている。これは、エネルギー獲得を優先しているといえる。従って、行動学習中でのタスク間の優先関係の変化に対応できていると判断する。

9. まとめ

本研究では、複数タスク下での行動選択について、各タスクの重要性の変化に注目し、各タスクの重要度に基づく行動選択手法を提案した。また、シミュレーション実験によって各タスクの重要度に応じた自律的な行動決定が可能であることを示した。

参考文献

- [1] 加藤 龍憲, 鈴木 昭二, 浅田 稔: “複数の報酬による強化学習を用いたサッカーロボットのゴール守備行動の獲得”, ロボティクスシンポジウム予稿集, Vol.4, pp.289-294, 1999
- [2] 上岡 拓未, 内部 英治, 銅谷 賢治: “複数の行動価値を用いた多目的強化学習” 電子情報学会信学技報, Vol.105, No.658, pp.127-132, 2006