センサ情報の自律的選択による効率的な行動選択の実現

室蘭工業大学 情報工学科 4 年　沼田利伸

2012 年 2 月 15 日

学習ロボットとセンサの問題点

近年では，多彩なセンサの開発によって，ロボットが適応可能な環境や実行可能なタスクが拡大した．また，1 台のロボットで複数のタスクを実行することの要求が高まっている．1 台のロボットが実行するタスクが増えれば，その分だけ多くのセンサを搭載しなければならない．そのため，近年では，多くのセンサを搭載したマルチタスク・ロボットが出現している．

これまで，ロボットに搭載するセンサは，人間の予測によって選定されていた．これは，ロボットが実行するタスクが簡単であるためや，ロボットが直面する環境が大きく変化しないために可能なことであった．

しかしながら，本研究では，ロボットがタスクを実行する上で必要となるセンサを特定することは難しいと考える．なぜなら，ロボットが直面する環境の変化によって，必要となるセンサも異なることが考えられるからである．また，人間にとっても未知・予測困難な環境へロボットを投入するとき，必要となるセンサを特定すらできないことが考えられるからである．

ここで，近年では，強化学習をはじめとする学習アルゴリズムをロボットに適用する事例が多くある．学習においても，その入力はセンサ情報[*1] であることがほとんどである．一般に，学習ではその入力次数が増えると，学習に要するコスト—メモリ，計算時間—も増大する傾向がある．そのため，このような問題を解決するために，ロボットが搭載する全てのセンサ情報を入力するのは望ましくないことが考えられる．

本研究の目的とアプローチ

本研究では，学習を適用したロボットが，センサ情報を自律的に選択し，効率的に行動選択する方法を実現することを目的とする．

このことを実現するために，センサ情報の必要・不要を自律的に判別する方法を考える．そのために，センサ情報の必要性を示す定量的な指標を考える．ここで，ロボットが環境の変化にも自律的に追従するために，環境を要因とするものを考える．

そこで，本研究では，環境の物理量とタスクの進捗度に相関があることに着目する．センサ情報は環境の物理量と対応するため，センサ情報とタスクの進捗度にも相関があると考えられる．

したがって，本研究では，タスクの進捗度とセンサ情報を相関分析することを考える．そして，そのときに得られる相関係数をセンサ情報の必要性として評価することで，必要・不要と判別できると考える．

提案手法

近年では，ロボットの学習として強化学習 [1] を採用する事例が多い．そこで，本研究では強化学習の枠組みの中で，センサ情報を自律的に選択する方法，選択したセンサ情報を利用する方法を考える．

はじめに，本研究が提案する学習機構の概念を図 1 に示す．この学習機構は，強化学習を基本とし，提案手法と連係して機能する．以下では，ロボットが n 個のセンサを搭載していることを前提として説明する．

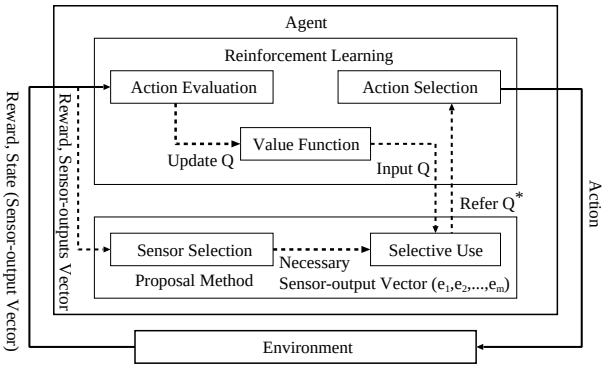


図 1　提案する学習機構の概念図

Sensor Selection では，センサ情報の選択を実現する．ここで，タスクの進捗度は，時刻 $t(t \geq 0)$ ごとに与えられる即時報酬 r_t によって表現する．このモジュールでは，現時刻 T までに記録した報酬値 r の集合 R_T と，n 個のセンサ情報 $e_1, e_2, \ldots, e_n$ の集合 $E_{1,T}, E_{2,T}, \ldots, E_{n,T}$ を相関分析する．そして，このときに得られた相関係数 $\rho_{1,T}, \rho_{2,T}, \ldots, \rho_{n,T}$ の絶対値が，任意に設定した閾値 ρ 以上であるとき，それに対応するセンサ情報を必要と判別する．

Selective Use では，選択したセンサ情報の利用を実現する．強化学習では，状態 s の定義にセンサ情報を用いることがある．ここで，センサ情報 E を状態変数 U に対応付ける写像 f があるとする．このとき，ある時刻 t の状態 s_t は次式の通りに定義される．

$$s_t = \{(u_{1,t}, u_{2,t}, \ldots u_{n,t}) | u_{i,t} = f(e_{i,t}), 1 \leq i \leq n\} \quad (1)$$

ここで，先述した方法で，$m(m \leq n)$ 個のセンサ情報の集合 $E_{1,t}, E_{2,t}, \ldots, E_{m,t}$ が必要，残りの $E_{m+1,t}, E_{m+2,t}, \ldots, E_{n,t}$ が不要と判別されたとする．このとき，先の状態を次式の通りに再定義する．

$$s_t^* = \{(u_{1,t}, u_{2,t}, \ldots u_{m,t}) | u_{i,t} = f(e_{i,t}), 1 \leq i \leq m\} \quad (2)$$

[*1] センサが計測した物理量は，電気的な量—電圧，電流—で表現される信号として出力される．この電気信号は，A/D コンバータで量子化され，離散値に変換される．本研究が定義するセンサ情報は，その変換で得られる離散値を指す．

ここで，不要なセンサ情報に対応する状態変数の集合の直積 $U_{m+1} \times U_{m+2} \times \ldots \times U_n$ を $\bar{U}$，その列 $(u_{m+1}, u_{m+2}, \ldots, u_n)$ を $\bar{u}$ とおく．このとき，このモジュールでは，状態 s_t^* における任意の行動 a の価値 $Q^*(s_t^*, a)$ を，次式の通りに計算する．

$$Q^*(s_t^*, a) = \frac{\sum_{\bar{u} \in \bar{U}} N(s_t^*, \bar{u}, a) \cdot Q(s_t^*, \bar{u}, a)}{\sum_{\bar{u} \in \bar{U}} N(s_t^*, \bar{u}, a)} \quad (3)$$

これは，$Q(s_t, a)$ の評価回数 $N(s_t, a)$ を重みとする加重平均である．提案手法を適用したエージェントが行動選択するときは，$Q^*(s_t^*, a)$ を参照するものとする．ただし，全てのセンサ情報を必要または不要と判別したとき，本来の $Q(s_t, a)$ を参照するものとする．

これにより，提案手法を適用したエージェントは，自律的に選択したセンサ情報による状態認識と，それによる効率的な行動選択が実現できると考えられる．

実環境における実機実験と考察

本実験では，提案手法の有効性を検証することを目的とする．そのために，一般的な強化学習と提案手法を別々の時に図 3(a) のロボットに適用し，図 2 のようなタスクを実行させる．このタスクでは，距離 d_A で目標を決定できるため，壁 A に正対する距離センサが必要，壁 B に正対する距離センサが不要となる．

このタスクの達成度は，壁 A からの距離 d_A の大きさに反比例する．そこで，時刻 t の距離 $d_{A,t}$ に反比例する状態変数を $u_{1,t}(1 \leq u_1 \leq 11)$ とおくとき，エージェントに与える即時報酬 r_t は次式の通りとする．

$$r_t = 11 - u_{1,t} \ (0 \leq r_t \leq 10) \quad (4)$$

エージェントは，図 3(b) の環境を等間隔の距離に分割して，大きさ 11×11 の状態空間として状態を認識するように設定する．エージェントが行動として前後左右の方向へ移動するときは，オムニホイール・モータを回転させる．オムニホイール・モータは，能動回転する方向に対して垂直な方向に受動回転できる．その他の実験設定の要約は，表 4.1 の通りである．

実験結果を図 4，図 5 に示す．図の通り，提案手法を適用した場合の方が，多くの試行において，より多くの報酬を安定して獲得していることが分かる．また，提案手法を適用したエージェントは，全試行において，壁 A に正対する距離センサによって状態認識し，それに基づいて行動選択していることが分かる．

これらのことから，提案手法を適用した場合の方が，より少ない状態行動対で学習できるため，タスクを安定して早くに達成できるのだと考えられる．

表 4.1　実験設定の要約

相関係数 ρ の閾値	0.8
行動選択手法	ϵ-greedy
探査的な行動の確率 ϵ	0.2
行動評価手法	Q-Learning
学習率 α	0.5
割引率 γ	0.5
行動価値 Q の初期値	0.0
選択可能な行動 A	{前進，右移動，後退，左移動，静止}
遷移可能な状態数 $\lvert S \rvert$	$121 (= 11 \times 11)$
初期状態 s_0	図 3(b) の左下の位置に相当する状態
1 試行	1000 回の行動選択
試行回数	30

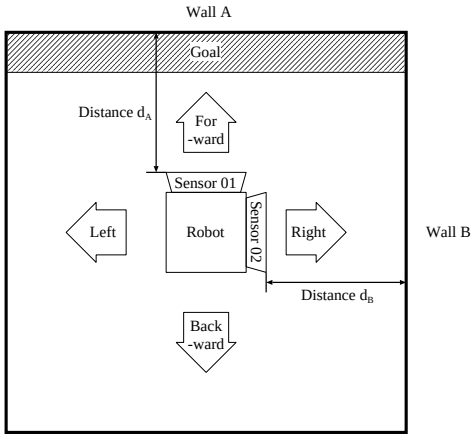
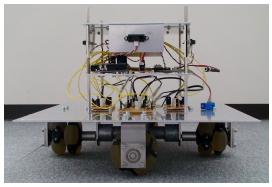
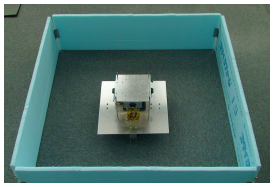


図 2　タスクの概要



(a) 実験機の移動ロボット



(b) 正方形状の実環境

図 3　実験機と環境

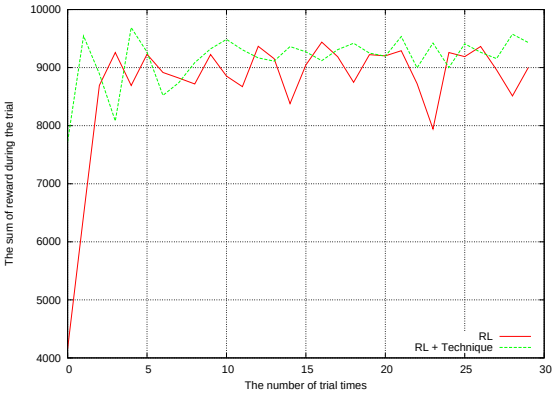


図 4　試行回数に対する累計報酬の推移

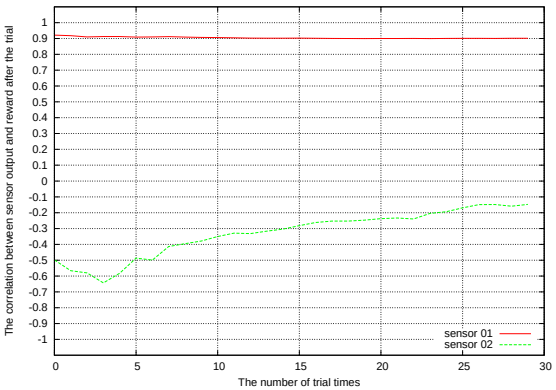


図 5　試行回数に対する相関係数の推移（提案手法）

今後の課題は，他の環境においても同様に検証し，より多くの環境で有効性を確立することが挙げられる．

参考文献

[1] Richard S. Sutton, Andrew G.Bart (訳：三上貞芳, 皆川雅章): 強化学習 森北出版株式会社 2001