経験情報に基づく不完全知覚の解決

― 確率的手法による知識量の抑制 ―

室蘭工業大学 情報工学科 4 年 認知ロボティクス研究室　平間経太

1. はじめに

近年，実ロボットに用いられることが多い学習法として強化学習がある．本研究では強化学習におけるロボットの状態認識に注目し，不完全知覚という問題を挙げる．強化学習を実ロボットへ適用する際，ロボットはセンサを用いて環境(以下，環境状態とする)を観測し，観測値から状態(以下，観測状態とする)を認識する．ロボットは各観測状態に対してタスクを達成出来る行動を学習していく．しかし複雑な環境では異なる状態を同じ状態であると認識してしまう不完全知覚という問題があり，タスクを達成出来る行動の学習が困難になる場合がある．従って，複雑な環境でも学習を行うために不完全知覚を解決する必要がある．

2. 先行研究

ロボット自身の経験情報を用いて不完全知覚を解決した先行研究[1]では，強化学習を適用したロボットが自身の経験情報とセンサ情報を用いて状態認識を行う．ロボットは自身の経験情報を利用して不完全知覚であると判断した状態を，不完全知覚である状態と，直前の状態と行動を用いて表す新しい状態に細分化して知識化する(図 2.1)．先行研究では，この不完全知覚である状態を直前の状態遷移で表した知識を状態知識と呼び，センサと合わせて状態認識に利用することで不完全知覚を改善し，学習が可能になる．

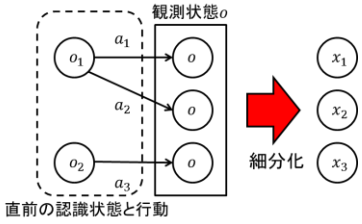

図 2.1. 不完全知覚である観測状態の細分化

3. 先行研究の問題点

先行研究では不完全知覚である観測状態を確定的に細分化している．しかし，不完全知覚はタスク達成に悪影響がない場合もある．そのため先行研究ではロボットが行動するにつれ無駄な細分化を行い，タスク達成には不要な状態知識を記憶していく．

4. 本研究の目的とアプローチ

本研究では，ロボットが自身の経験情報に基づいて確率的に細分化を行い，知識量を抑制しつつ不完全知覚を改善する手法を提案する．そのために本研究では不完全知覚とタスクの関係に注目する．不完全知覚が起きていてタスクの達成に悪影響がある場合には高い確率で，タスクへの悪影響が少ないと考えられる場合には低い確率で細分化を行うことで，不要な細分化を防ぐ．

5. 提案手法

提案手法と強化学習の関係を図 5.1 に示す．提案手法は強化学習とは独立に働く．ただし，本手法では細分化によりロボットが認識出来る状態が増加するため，それに伴い強化学習の学習空間を拡大する必要がある．

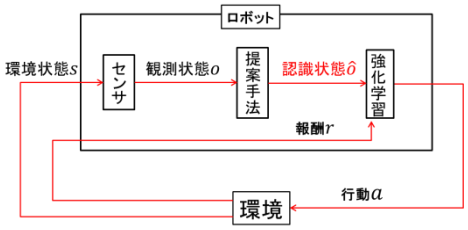

図 5.1　提案手法と強化学習の関係

本研究の提案手法を適用したロボットは，以下の 3 つの経験情報を，自身が経験した状態遷移の回数を基に算出する．

・ある状態と行動によって遷移する状態の分散
・ある状態で選択する行動の分散
・ある状態の経験回数

これらを用いてある状態遷移$(\hat{o}_i, a_j, \hat{o}_k)$を経験したときに細分化を行う確率を式(5.1)で決定する．式(5.1)において，$\sigma_{\hat{o}}{}^2$は遷移する状態の分散，$\sigma_a{}^2$は選択する行動の分散，$\mathrm{N}(\hat{o}_i, a_j, \hat{o}_k)$はこの状態遷移の経験回数，$\theta$ は各閾値を表している．また，ς_bはゲインbのシグモイド関数である．この確率に基づいて細分化を行うことで，知識量が急激に増加することなく，必要な知識のみが作成される．

$$\begin{aligned} \mathrm{P}_{\mathrm{segmente}}(\hat{o}_i, a_j, \hat{o}_k) &= \varsigma_{b_N}(\mathrm{N}(\hat{o}_i) - \theta_N) \\ &\cdot \varsigma_{b_a}\left(\sigma_{\mathrm{a}}{}^2(\hat{o}_i) - \theta_a\right) \\ &\cdot \varsigma_{b_{\hat{o}}}\left(\sigma_{\hat{o}}{}^2(\hat{o}_i, a_j) - \theta_{\hat{o}}\right) \end{aligned} \qquad (5.1)$$

6. 実験

今回提案する手法の有効性を検証するためにシミュレーション実験を行った．実験により，提案手法で知識量が抑制出来ているか，不完全知覚を改善出来ているかを検証する．実験では状態認識の異なる 4 体のエー

ジェントに強化学習を適用し，迷路問題を行った．行った検証実験の概要を図 6.1 に，実験に用いた環境を図 6.2 に，各パラメータを表 6.1 に示す．また，実験結果として，各エージェントの各試行における行動数の推移を図 6.3 に，図 6.3 について先行研究と提案手法の結果を行動回数 100 までの範囲に拡大したものを図 6.4 に，先行研究と提案手法の状態知識の数の推移を図 6.5 にそれぞれ示す．

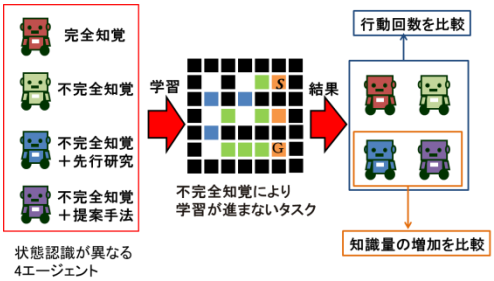

図 6.1　実験概要

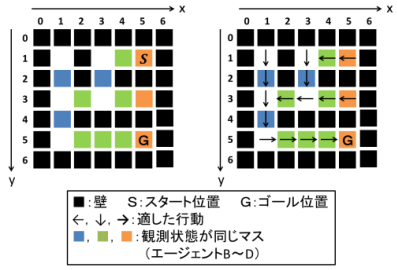

図 6.2. 不完全知覚で学習が行えないタスク

表 6.1　実験パラメータ

報酬(ゴール到達時のみ)	100
全試行数	500
スタート位置	(x,y)=(5,1)
ゴール位置	(x,y)=(5,5)
学習手法	Q学習
行動選択手法	ε −greedy
Q値の初期値	0.01
α（Q学習）	0.5
γ（Q学習）	0.7
ε（ε −greedy）	0.05
経験回数のシグモイドb_N, θ_N	$b_N = 0.3,\ \theta_N = 100$
選択する行動の分散のシグモイドb_α, θ_α	$b_\alpha = 30,\ \theta_\alpha = 1.0$
遷移する状態の分散のシグモイド$b_{\hat{\theta}}$, $\theta_{\hat{\theta}}$	$b_{\hat{\theta}} = 750,\ \theta_{\hat{\theta}} = 0.04$

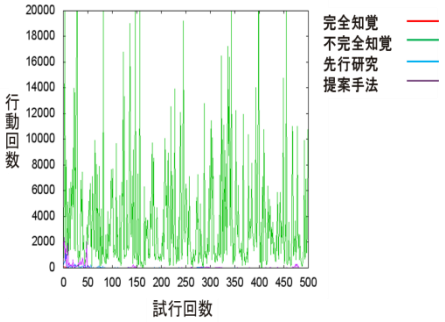

図 6.3　4 体のエージェントの各試行における行動回数の推移

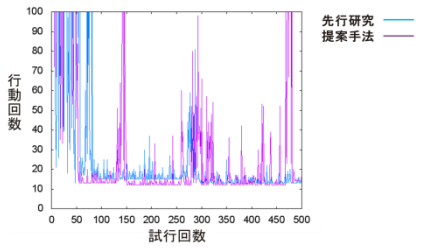

図 6.4　図 6.3 を行動回数 100 までの範囲で拡大．

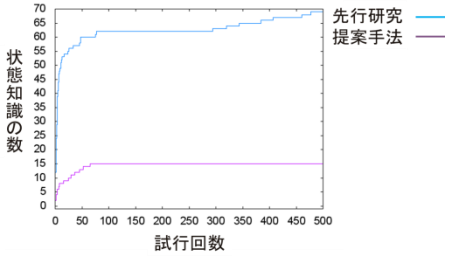

図 6.5　状態知識の数の推移

7. まとめ

シミュレーション実験の結果から，提案手法により，知識量を抑制しつつ不完全知覚を改善し，学習が可能になることを示した．

参考文献

[1] Yoshiki Miyazaki , Kentaro Kurashige: Estimate of current state based on experience in POMDP for Reinforcement Learning, Proceedings of the seventeenth International Symposium on Artificial Life and Robotics (AROB 17th '12) , pp.1135-1138 , (Jan. 19-21, 2012)