強化学習における不完全知覚の解決

― 経験情報に基づく状態認識 ―

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

　ロボットはセンサを用いて自身や周囲の状態といった環境を観測する．環境を観測する際に生じる問題として，不完全知覚と呼ばれる問題がある．不完全知覚とは，センサの種類や能力の不足により，本来異なる複数の状態をロボットが区別できず，同じ状態であると観測してしまう問題である．

本研究では，強化学習におけるロボットの不完全知覚について考える．

2. 強化学習における不完全知覚問題

　強化学習では観測した各状態に対して，それぞれ一つの適切な行動を学習する．しかし，不完全知覚が起きて，本来異なる複数の状態を同じ状態であると観測してしまった場合，適切な行動が学習できなくなるという問題がある(図2)．図2において$ s $は本来の状態，$ o $はエージェントが観測した状態，$ a $はその状態における適切な行動を表している．図2においてエージェントは，適切な行動が異なる状態 $s\_{1}$と$s\_{2}$を，同じ観測$ o\_{1}$として観測しているため，適切な行動は$a\_{1}$と$ a\_{2}$の二つ存在する．



図 1．強化学習における不完全知覚問題

具体例として，図3に迷路探索における例を挙げる．エージェントは周囲の壁の有無によって状態を観測する．このため，同じ色で示したマスで同じ観測を得る.この例において， のマスで観測される状態で適切な行動は，上部のマスでは左，下部のマスでは右への移動である．一つの観測に対して複数の適切な行動が存在するため，適した行動が右と左のどちらへの移動なのか学習できない．



図 2．迷路探索における例

3. 研究目的

　本研究では，不完全知覚が起きている観測を分割する手法を提案し，強化学習における不完全知覚の解決することを目的とする．

4. アプローチ

　本研究では，不完全知覚により適切な行動が複数存在する観測を，分割することを考える．これは図2より，適切な行動が複数ある状態では，本来異なる状態を同じ状態として観測していると考えられるためである．つまり，本来の状態に分割することが出来ればよいと考えられる．

以上のことから，まず観測した状態で適切な行動が複数あるか判別する「観測判別部」と，適切な行動が複数ある場合に観測を分割する「観測分割部」が必要だと考えられる．

　これまでの成果として，観測した状態に適切な行動が複数あるか判別を行う方法を考察し，検証実験を行った．

4.1 分割する観測の判別

適した行動が複数ある観測における行動，状態遷移に以下の特徴が考えられる．

・選択する行動が一つに確定しない

・同じ行動を選択しても，遷移する状態が一つに確定しない

よって，分割が必要な観測において，以下の二つの確率にばらつきが生じると考えられる．

・各行動を選択する確率

・ある行動をとった際に各状態へ遷移する確率

これら二つの確率に注目し，観測を分割する必要があるかを判別することを考えた．

5. 予備実験

5.1 実験目的

　注目する二つの確率を計算し，アプローチで述べたようにばらつきがあるかを確認することで，観測の判別に用いることが出来るか検証を行う．

5.2 実験概要

　不完全知覚が起こる強化学習エージェントを用いて，迷路探索を行わせた．エージェントは四近傍の壁の有無で状態を観測するため，不完全知覚が起きる．タスクはすべての観測で行動が一意に決まるタスク1と，行動が一意に決まらない観測を含むタスク2である．図4に実験環境を示す．図4において，同じ色のマスは同じ観測が得られるため，不完全知覚が起きる．タスク2において観測$o\_{1}$が得られるマスにおいて，適した行動が右への移動と，左への移動の二つある．



図 3．実験環境

本実験では二つの確率の計算のため，行動の選択確率を式(1)，状態の遷移確率を式(2)で計算した．

$$P\left(o\_{t},a\_{t}\right)=\frac{N(o\_{t},a\_{t})}{N(o\_{t})}　　　　　　　　(1)$$

$$P\left(o\_{t},a\_{t},o\_{t+1}\right)=\frac{N(o\_{t},a\_{t},o\_{t+1})}{N(o\_{t},a\_{t})}　　　　(2)$$

$N(o\_{t})$:観測$o\_{t}$を得た回数

$N(o\_{t} , a\_{t} )$:状態行動対$(o\_{t} , a\_{t})$を経験した回数

$N(o\_{t} , a\_{t} , o\_{t+1} )$:状態遷移$(o\_{t} , a\_{t} ,o\_{t+1} )$を経験した回数

実験パラメータを表1に示す．

表 1．実験パラメータ



5.3 実験結果

　まず，各行動の選択確率を示す．



図 4．各試行における各行動の選択確率の推移

図4の左図から，適切な行動が一つである観測では行動の選択確率にばらつきがないことが確認できた．また，右図から，適切な行動が複数ある場合，行動の選択確率にばらつきがあることが確認できた．

　次に各状態への遷移確率を図5に示す．図5の左図から，適切な行動が一つである観測では遷移先は一つで確率は1.0になることが確認できた．また，右図から，不完全知覚が起きている観測では遷移先が複数あり，遷移確率が0にならないことが確認できた．



図 5．各試行における各状態への遷移確率

5.3 実験まとめ

　予備実験の結果より，各状態への遷移確率から，少なくてもその観測で不完全知覚が起きているか判別できると考えられる．また，各行動の選択確率から，少なくても環境内に行動が決まらない観測が存在しているか判別できると考えられる．

6. 今後の予定

　・観測判別方法の具体化

　　　これまでの成果から，観測を分割する際の情報に二つの確率情報を用いることが出来ると考えられる．今後は二つの確率情報をどのように利用するか決め，検証を行う．

・観測分割方法の検討，提案手法の提案

　　観測分割手法の検討と検証を行う．観測判別手法と合わせて実装し，提案手法の検証，提案を行う．

　・観測判別手法の提案

　　　これまでのペースから，観測の分割手法に　まで手が回らない可能性がある．その場合，観測の分割の部分を除いた観測判別手法の提案を行う．本研究室で提案された状態認識手法[1]を利用し，検証を行う．

参考文献

[1] 宮崎愛央,不完全知覚に対する状態認識法

の提案-経験情報に基づく現状態の推定-,

修士研究論文,2012.