

マルチエージェント探索問題における粗視化とフィルタリングの統合手法による領域分割について

湯徳尊久[†] 杉山歩未[†] 菅原俊治[†]

[†] 早稲田大学

1 はじめに

ロボットは産業ロボットだけでなく日常面でも身近なものとなっており、清掃、警備などの様々な場面において重要度が増している。[1]は、清掃問題においてロボットの視野を環境の状態としたQ学習を用いることでロボットが協調行動を獲得する手法を提案している。しかし、実際の場面では状態空間が非常に大きくなり学習も困難となる。そのため協調行動を取ることが難しくなる。状態空間を削減する手法は様々なものが存在するが、複雑な環境での高度な協調が求められる状況において、局所的な視野しか持たないエージェントによって効率的な探索を実現するための手法は十分に考慮されていない。

本研究では、マルチエージェント探索問題において、処理に要求されるエージェント数の異なる複数種類の対象物がある環境を想定する。このような環境では学習のための状態空間が爆発的に増加し、既存手法では学習が極めて困難となる。そこで我々は、状態空間を削減するための手法として粗視化とフィルタリングの2つに注目し、[2]にて、視野から得られる情報をフィルタリングにより制限し、さらに粗視化を用いて状態空間を抽象化するという2手法を統合し、それぞれの手法の欠点を補うことで学習の効率を向上させる手法を提案した。しかし、大規模な領域での有効性は確認していなかったため、一般的に高い効率を実現する領域分割手法に提案手法を適応した場合と比較し、局所的な視野しか持たないエージェントで構成されるシステムが、大規模な空間に適応するための手法を検討する。

2 問題設定

環境は2次元格子状とし、エージェントや対象物は各マスに存在する。1つのマスには対象物は最大で1個、エージェントは複数存在できるとした。時間は離散時間とし、最小単位は1ステップとする。1ステップでエージェントは移動と対象物の回収を同時に行う。対象物の集合を $\mathcal{T} = \{T_1, \dots, T_m\}$ とする。各対象物は $T_l = \{c_l, r_l\}$ で表され、 c_l は対象物の種類、 r_l はエージェントが T_l の回収に成功したときに得られる報酬で

ある。本研究では回収に複数体のエージェントが必要な対象物を想定する。回収に必要なエージェント数は c_l によって決まる。

エージェントの集合を $A = \{1, \dots, n\}$ とする。エージェント i は視野 N_{view}^i (0以上の自然数) をもち、 N_{view}^i ステップで移動できるマスまでを視野とし、エージェントと対象物を認識できる。また、エージェント間の通信は行わない。本モデルでは、エージェントの行動は上下左右いずれかのマスへの移動と現在のマスへの滞在とする。

本研究では、エージェントが全ての対象物を回収するまでを1エピソードとし、1エピソードにかかったステップ数を回収効率の評価指標とする。

3 提案手法

3.1 視野制限手法(フィルタリング)

視野範囲内にある対象物とエージェントに対して、観測エージェントから近い k_t 個の対象物とその対象物から近い k_a 体のエージェントを状態表現として記憶する。ただし、 k_t, k_a は非負の整数である。なお、状態表現について、観測エージェントから確認できる観測物の相対座標を認識し、それらを(相対座標, 観測物を示す記号)の組として順に全て羅列した文字と数字の列を記録し、視野状態として学習に使用する。ここで、観測物はエージェントか対象物を表す。

3.2 視野の粗視化

粗視化手法では、エージェント i から N_{near} (ただし $N_{view}^i \geq N_{near}$) ステップの範囲の視野を新たに定義し、この範囲内はエージェント、対象物共に詳細な座標まで観測するが、その範囲外は対象物のある方向だけを観測し、最後に学習に使用する状態として、エージェントの総数を登録する手法である。

3.3 フィルタリングと粗視化の統合

3.1節で述べたフィルタリングと3.2節で述べた粗視化手法を組み合わせる。つまり、予めフィルタリングによってエージェントの視野から得た情報を、エージェント i から近い k_t 個の対象物とその対象物から近い k_a 体のエージェントまで制限し、その状態で粗視化手法を用いて強化学習の状態とする。

題名

Takahisa Yutoku[†] Ayumi Sugiyama[†], Toshiharu Sugawara[†]

[†]Department of Computer Science and Engineering, Waseda University

	パラメータ	値
視野範囲	N_{view}	5,15
エージェント数	$ A $	40
$c_l = 1$ の対象物の数	N_{c1}	10
$c_l = 2$ の対象物の数	N_{c2}	10
$c_l = 3$ の対象物の数	N_{c3}	10
報酬	r	10
学習率	α	0.1
割引率	γ	0.9
	ε	0.1

表 1: 各パラメータの名称と値

4 実験

4.1 実験環境

エージェントが探索する領域を 50×50 の 2 次元グリッドとし、便宜上この領域を 25×25 の 4 個のサブ領域に区切る。本稿では、エージェントの担当領域について、2 パターンの環境で実験を行う。環境 1 では、エージェントは領域分割をせず、全領域を担当する。環境 2 では、エージェントは 4 つのサブ領域に均等な台数で配置され、配置されたサブ領域をそのエージェントの担当領域とする。エージェントは自分の担当領域外を探索せず、担当領域外に存在するエージェントと対象物については認知せず、学習の状態表現に含めない。両方の環境において回収対象物はランダムに配置する。

それぞれの環境について、視野範囲を 5 ステップ、15 ステップ先まで観測可能にした場合でその影響を調査した。表 1 に実験に用いるパラメータの値を示す。また、本実験では、3.1 節のフィルタリングの各パラメータとして、 $k_t = 1, k_a = 3$ と $k_t = 3, k_a = 3$ の結果を示す。便宜上、それぞれの場合を **1T3A**、**3T3A** と呼称する。各パラメータで N_{near} の値を変化させ、その違いを調査した。本実験では回収時の報酬 r は対象物によらず同じ値が配分する。実験は 3000 エピソードの学習を 1 試行とし、結果には 15 回の試行の平均値を用いた。

4.2 実験結果と考察

環境 1、環境 2 の実験について、視野範囲が 5 の場合を表 2 に、視野範囲が 15 の場合を表 3 に示す。なお、表中の平均ステップ数とは学習が安定した 2000 から 3000 エピソードにおける平均値を用いた。どちらの表からも、領域分割のない環境 1 のほうが効果的に対象物の回収に成功したことがわかる。原因は 3 つ挙げられる。第 1 に対象物の配置にサブ領域で偏りができたエピソードにおいて、複数体で回収可能な対象物が多数存在し、各エージェントが 1 つの対象物に留まり他のエージェントの到着を待つ、停止状態ができたことである。第 2 に領域分割によって、エージェントが回収しにくい探索領域の四隅が増加したことである。第 3 には領域分割ありの場合、自分の担当領域の対象物

名称	平均ステップ数		状態数	
	環境 1	環境 2	環境 1	環境 2
1T3A, $N_{near} = 1$	580	707	726	679
1T3A, $N_{near} = 2$	534	642	4862	4591
1T3A, $N_{near} = 3$	521	629	21754	20901
3T3A, $N_{near} = 1$	586	710	11558	10438
3T3A, $N_{near} = 2$	579	693	52628	50025
3T3A, $N_{near} = 3$	682	822	164082	166188

表 2: 各環境での平均ステップ数と状態数 (視野範囲 5)

名称	平均ステップ数		状態数	
	環境 1	環境 2	環境 1	環境 2
1T3A, $N_{near} = 1$	118	149	1077	1041
1T3A, $N_{near} = 2$	112	135	8112	7875
1T3A, $N_{near} = 3$	125	144	29513	29132
3T3A, $N_{near} = 1$	280	331	47881	42393
3T3A, $N_{near} = 2$	419	600	229747	221080
3T3A, $N_{near} = 3$	710	1425	50 万以上	50 万以上

表 3: 各環境での平均ステップ数と状態数 (視野範囲 15)

を早めに回収できたとしても他の担当領域に入れないため、以降その担当領域に存在するエージェントを有効に働かせることができない。また、視野範囲 5 と 15 の場合を比較すると、ほとんどの場合で視野範囲 15 のほうが回収効率が良いことがわかる。したがって広範囲において観測可能な状況では、最適解に近い行動が学習可能であると推測できる。しかし、広範囲を観測可能とすると、視野内にいる大量のエージェントや対象物も観測可能となり、フィルタリングによる情報の遮断を若干緩めただけで、たちまち状態数が爆発的に増加し、学習が困難な状況が存在しやすくなる。また、学習が困難である状況で、領域分割を行うと、前述の原因が顕著に現れ、回収効率の悪化にもつながると思われる。

5 まとめ

本研究は、マルチエージェント探索問題において、処理に要求されるエージェント数の異なる複数種類の対象物のある環境を想定し、我々の提案した粗視化とフィルタリングの 2 手法の統合手法による大規模領域での有効性を調査した。提案手法を適用した領域分割は欠点も多く、領域分割をせず提案手法を適用した場合のほうが、効果的に対象物の回収を行うことを確認した。

参考文献

- [1] Mengchun Xie. Representation of the perceived environment and acquisition of behavior rule for multi-agent systems by q-learning. In *Proc. of 4th Int. Conf. on Autonomous Robots and Agents 2009*, pp. 453–457, IEEE.
- [2] 湯徳尊久, 杉山歩未, 菅原俊治. 探索問題におけるフィルタリングと粗視化を用いたマルチエージェント強化学習について. 電子情報通信学会技術研究報告 (人工知能と知識処理研究会), Vol. 116, No. 350, pp. 55–60, 2016.