環境の変動に適応する移動ロボットの行動獲得

○港 隆史　　浅田 稔
大阪大学大学院 工学研究科

Environmental Change Adaptation for Mobile Robot Navigation

○Takashi Minato　　Minoru Asada
Osaka University

Abstract — *This paper proposes a method which adapts robots to environmental changes by transfering a learned policy in the previous environments into a new one and modifying it to cope with these changes. We apply the method to a mobile robot navigation problem of which task is to reach the target avoiding obstacles based on sonar and visual information.*

1　はじめに

種々の環境で作業する自律ロボットを実現するためには，ロボットの環境の変動への適応過程が重要となる．従来研究では，環境ごとに行動政策あるいは制御コントローラを持つことにより環境の変動に対処していた．このような方法では

1. 新たな未知の環境ではそこに関する知識を持たないために学習時間がかかる．
2. 環境ごとに異なる行動政策を持つため，多大な記憶領域を必要とする．

という問題がある．

1. の問題に対して，Thrun et al. [2, 3] は Lifelong Learning, また Tanaka and Yamamura[1] は Lifelong Reinforcement Learning という枠組みを提案している．これらの手法ではあるクラスの環境を想定し，そのクラス内の新たな環境にロボットが遭遇すると，それまでに得られた環境に関する不変な知識を現在の学習に利用することにより学習時間を短縮させている．しかし，経験してきた環境で作業するためには環境ごとに行動政策を持たなければならず，上述した 2. の問題が考慮されていない．

本報告では，環境の変動に応じて単一の行動政策を部分的に修正する手法を提案する．ロボットは環境の変動を認識すると，行動政策の中でタスク達成に不都合が生じる部分だけを局所的に修正する．大部分の行動政策を残しておくことにより，学習時間が短縮され，さらに環境の変動を吸収した単一の行動政策が獲得される．提案した手法を，障害物を回避し目標物に到達する移動ロボットのナビゲーション問題に適用し，シミュレーション結果および実験結果を示す．

2　環境の変動への適応

強化学習で得られる行動政策 (状態集合$\boldsymbol{S}$から行動集合$\boldsymbol{A}$への写像) は環境の状態遷移の情報を含んでいる．そのため，環境の変動に伴い状態遷移が変化すると行動政策は新たな環境で最適政策を示さない．しかし最適性は失われるが部分的に適用できる可能性はある．そこで本手法では，環境の変動をロボットが認識したとき，タスク達成に不都合な状態を探索し，それらの状態の行動政策を学習により修正する．ここでロボットは，環境の変動を自身のタスク成功率の低下によって認識する．

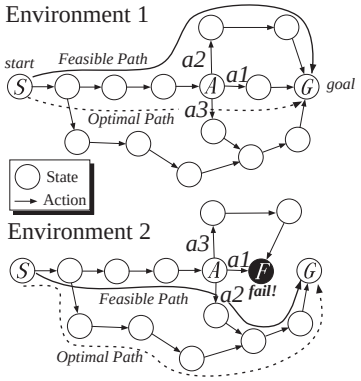


Fig.1 Policy modification

このことを状態遷移が異なる 2 つの環境の例 (Fig.1) を用いて具体的に説明する．環境 1 で獲得された行動政策 (optimal path) を環境 2 に適用すると，ロボットは状態 A でタスクに失敗する．そこでロボットは状態 A の行動政策を学習により修正し，$A \rightarrow \boldsymbol{a}_2$なる新たな行動政策を獲得する．この行動政策は，最適ではないがどちらの環境でもタスクを達成することができる．このようにして環境の変動に適応する．以上をアルゴリズムにしたものを以下に示す．学習には Q 学習を用いる．

1. 初期環境でタスク成功率 RS が決められた値を越えるまで Q 学習を行う．得られた行動政策を $P : \boldsymbol{S} \rightarrow \boldsymbol{A}$とする．
2. RS が低下しない限り P を使い続ける．低下すれば 3 へ．
3. 行動政策を修正する状態集合$\boldsymbol{S}_r \subset \boldsymbol{S}$を探索し，その環境での RS がβにより決まる成功率に回復するまで P を修正する．学習中の状態$\boldsymbol{s}$における行動戦略は

   $\boldsymbol{s} \notin \boldsymbol{S}_r \bigcup \boldsymbol{S}_n$ なら P

   それ以外なら通常の Q 学習の行動戦略

   に従う．ここで$\boldsymbol{S}_n \subset \boldsymbol{S}$は未経験状態集合である．
4. 得られた行動政策を P として 2 へ．

ここでβは適応率であり，環境の変動時に低下した成功率をどれだけ回復するかを示す割合である．βは次のように定義される．

$$RS_d = RS_c + \beta(RS_p - RS_c) \quad (1)$$

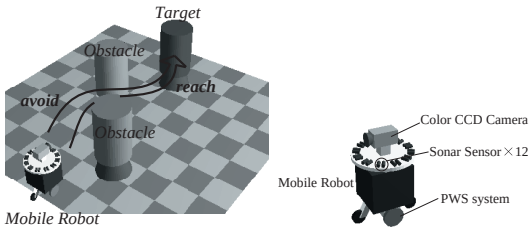
ここでRS_p, RS_c, RS_dはそれぞれ前の環境での成功率，現在の環境での成功率，目標成功率である．

3 実験結果

提案した手法を，静止障害物が存在する環境下でそれらを回避しながら指定された静止目標物に近づく移動ロボットのナビゲーションタスク (Fig.2(a)) に適用した．環境の変動として障害物の配置・個数の変化を設定し，βは 0.5 とした．

報酬はロボットが目標物に到達したときに与えた．また行動集合は左右輪の前進・停止・後退の組み合わせ，状態集合は次の 4 つの状態変数からなる空間を適当に離散化したものとした (全状態数 3060)．

- 画像上の目標物の水平位置と大きさ
- 最小測定値を示す超音波センサの方向とその値



(a) task (b) robot

Fig.2 Task and mobile robot

環境として Fig.3 に示す 5 つを用意した．図中の黒円が目標物，それ以外の円が障害物，多角形は移動ロボットの軌跡である．初期環境を $E1$ とし，$E2$ から $E5$ へと順に適応させた．各環境で獲得された行動政策を用いたときのタスク成功率を Table1 に示す．i 番目の環境 Ei で獲得された行動政策を Pi としており，成功率は Pi を Ei に適用したときのものを示している．また$\boldsymbol{S}_r$ は Pi を獲得する際に修正された状態数である．

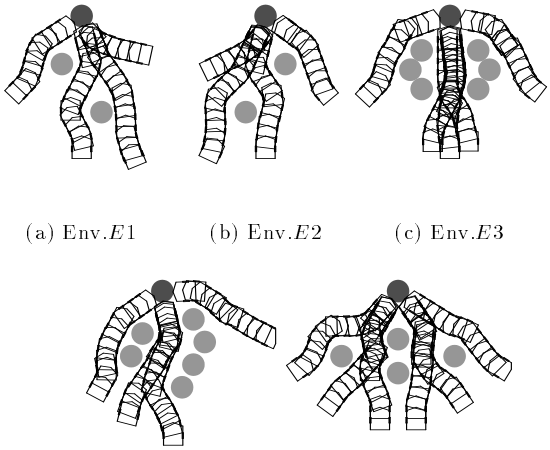
Table1 から分かるように，ロボットの置かれた環境が未経験の環境に変動するとタスク成功率が低下するが，新たな環境で学習しなおすことにより，その環境およびそれまでに経験した環境でタスクを達成できる行動政策を獲得している．最終的にロボットは 5 つの環境でタスクを達成できる 1 つの行動政策 $P5$ を獲得しており，本手法の有効性が示された．Fig.3 のロボットの軌跡は $P5$ を各環境に適用した結果である．

本手法を用いると学習時間は短縮されるが，最適性は失われる．たとえば Fig.3 の $E5$ における最も左の軌跡は障害物を回避したあと右に遠回りしており，最適性が失われている．

最後に $P5$ を実機に搭載した結果を Fig.4 に示す．

Table 1 Success rates of each policy [%]

policy	$\boldsymbol{S}_r$	$E1$	$E2$	$E3$	$E4$	$E5$
$P1$	-	90.9	61.0	(48.2)	(47.3)	(61.5)
$P2$	61	92.4	90.8	45.2	(50.2)	(69.5)
$P3$	75	88.2	88.8	93.3	82.7	(68.9)
$P4$	44	82.2	87.0	93.2	87.5	67.4
$P5$	69	89.4	89.9	89.8	76.4	84.7



(a) Env.$E1$ (b) Env.$E2$ (c) Env.$E3$

(d) Env.$E4$ (e) Env.$E5$

Fig.3 Environments and successful trajectories

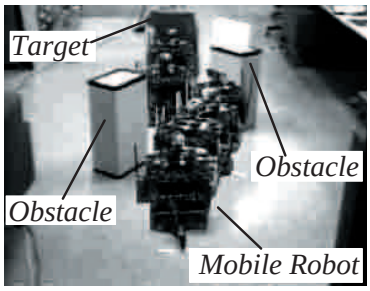


Fig.4 Experimental result

4 考察

本手法では，環境間で政策が競合する状態の政策は，後から学習された結果が優先される．その結果，過去の環境での成功率が低下する可能性がある．本手法の課題は，そのような状態をできる限り少なくするための環境の出現順序，あるいは環境のクラスの決定方法の考案である．

参考文献

[1] F. Tanaka and M. Yamamura. An approach to lifelong reinforcement learning through multiple environments. In *Proceedings of Sixth European Workshop on Learning Robots*, pp. 93–99, 1997.

[2] S. Thrun. A lifelong learning perspective for mobile robot control. In *Proceedings of IEEE/RSJ/GI International Conference on Intelligent Robots and Systems*, Vol. 1, pp. 23–30, 1994.

[3] S. Thrun and T. M. Mitchell. Lifelong robot learning. Technical Report IAI-TR-93-7, University of Bonn, Dept.of Computer SicenceIII, July 1993.