自律移動ロボットにおける時系列情報の導入による知覚騙し問題の解決

菊池司 羽倉淳 藤田ハミド
岩手県立大学 ソフトウェア情報学部

1. はじめに

未知の環境において,移動ロボットの行動様式を設計段階で構築しておくことは，設計者による環境の細部の予測が不可能に近いという理由により，非常に難しい.一般に，このような環境下で行動獲得や行動計画を自律的に行うことが効果的であるため,強化学習やニューラルネットワークといった学習手法を利用して行動様式を構築することが多い[1][2].これらの研究では，センサの観測から環境を同定し,その環境に適した行動選択を行うことが多い.しかし,実際には学習手法の扱うべき状態数を考慮し,センサの解像度を限定して利用する場合や実世界におけるセンサノイズの影響で知覚騙しが発生する.そのため,環境に応じた望ましい行動を行うことが困難となる.

以上の問題点を考慮し,本稿では,強化学習における状態空間として,現在の観測値のクラスタリングを状態とすることに加え,これらの状態の時系列を考慮した手法を提案する.本手法を利用することで,ロボットの経路情報が環境認識に用いられるため,異なる環境の混同が減少することが考えられる.その結果,部分観測マルコフ決定過程(POMDP:Partially Observable Markov Decision)としてモデル化される環境を,収束が保障されているマルコフ決定過程(MDP:Markov Decision Process)に近づけることが可能になるため,より適切な経路選択が可能になることが期待できる.

2. 知覚騙し問題

自律移動ロボットにおける環境の同定は,エージェントの持つ局所的なセンサによって行われる.しかし,このような観測の結果,センシングレンジの限定やセンサの観測精度の問題から図１に示すように,異なった環境として認識することを求められる環境を同一環境として誤認することがある.図１のAとBを比較したとき,エージェントの観測結果として同一の観測値が得られ,獲得される行動も同一の行動となる.しかし両者の置かれている環境は,俯瞰的な視点では異なる環境であり,行動とそれにより獲得されるべき評価も異なる.このような問題は知覚騙し問題として知られており,知覚騙し問題の発生する環境における行動決定過程は,POMDPとしてモデル化される.

しかし,知覚騙しの発生する要因がセンサ能力に起因するとは一概には言えない.人間をはじめとする多くの生物は,視覚と駆動系を持つ点でエージェントと類似しているが,知覚騙しが生じることがない.多くの場合,ロボットにおける駆動系の制御の信号は,センサから与えられた観測情報のみである.それに対し,生物における駆動系の制御は,視覚情報に加え,経験や予測といった記憶的要素も含めた制御となることが考えられる.つまり(1)センサの観測能力の限界,(2)記憶的要素の非考慮の 2 点が知覚騙しの生じる大きな要因であると考えられる.

以上のことから,状況判断を正確に行うために，この２点を考慮する必要があり，本稿では特に(2)の記憶的要素を考慮するために時系列情報を利用することを行う.

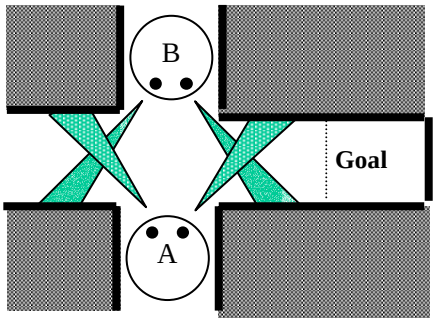

図1 知覚騙し問題

3. 状態の時系列を用いた強化学習法

従来,強化学習における状態空間の構築は,観測 θ を状態 s とする場合が多い[1][2][3].しかし,この場合，前述の問題から知覚騙し問題が発生することが多い.一方本手法では,離散時間 t における観測 θ_t を写像 ϕ により有限集合である観測状態 o_t へマッピングする.同様に過去に得られた離散時間 t-1 から t-i の観測状態 o_t から o_{t-i} までの集合を基に強化学習で用いる状態 s_t を構成する[3].以下に処理の流れを示す.

A Solution for Perceptual Aliasing Problem on Autonomous Mobile Robot by Using State Time Series
Tsukasa KIKUCHI ,Jun HAKURA ,Hamid FUJITA
Faculty of Software and Information Science ,Iwate Prefectural University

$$o_t = \Phi(\theta_t) \quad (1)$$

$$s_t \equiv \langle o_t, o_{t-1}, \ldots, o_{t-i} \rangle \quad (2)$$

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha \left[r + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \right] \quad (3)$$

3.1) 観測状態の獲得手法)

ここで,) 観測状態を構成する際の問題点として,観測誤差の存在が挙げられる.) そのため,) 観測誤差の処理を含むために,) 観測値を観測状態とすることは環境の同定が不正確になる可能性がある.) ここではこの問題を回避するために,) 写像ϕを導入しており,) 本稿では写像ϕを行うための手法として適応共鳴理論（ART：Adaptive) Resonance) Theory）を用いる.) ART では観測値をパターンとして捉え,) 類似度に基づいて観測値を分類するため,誤差が状態の構成に与える影響が減少することが期待できる[2][5].)

4) 計算機シミュレーション)

環境は 8×8 のグリッドワールドで表し,) ロボットは斜めを含む 8 方向への移動を与える.) ロボットの観測する情報は,周囲 4 方向（前後左右）の障害物を 0 と 1 のバイナリ情報とし,) 衝突時に負の報酬を,) 目的達成時に正の報酬を与えた.) 開始地点を図 3 中の S（スタート）とし,) G（ゴール）へ向かうことを目的として与え,) スタート・ゴール間に障害物を設置することで,) 擬似的に PODMP 環境とし,) 提案手法の有効性を検証した.)

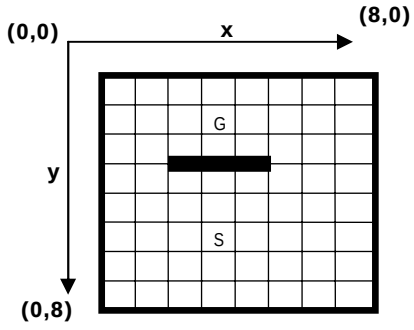

図2 実験環境

5) シミュレーション)

PODMP 環境下において,) 解が得られたことを図 3 に示す.) 1000 回の試行の中で,) およそ 100 回程度の試行回数で収束し,) ゴールまでの準最適解（準最短経路）を得ることができた.) しかし,) 実際にはゴールまでの最短 Step 数が 4 であることに対し,提案手法では,) 最も収束した値でも Step 数が 6 となった.) 提案手法によってマルコフ性が仮定されていると考えられる環境において Q-Learning を行った結果,) 最適解まで収束しないという結果に至った.) 時系列から得た観測情報を使用した場合,) 異なる場所の同一視はなくなったが,) 必要以上に状態数が増加したために同一の場所を異なる場所として捉えてしまう問題が実験により示された.)

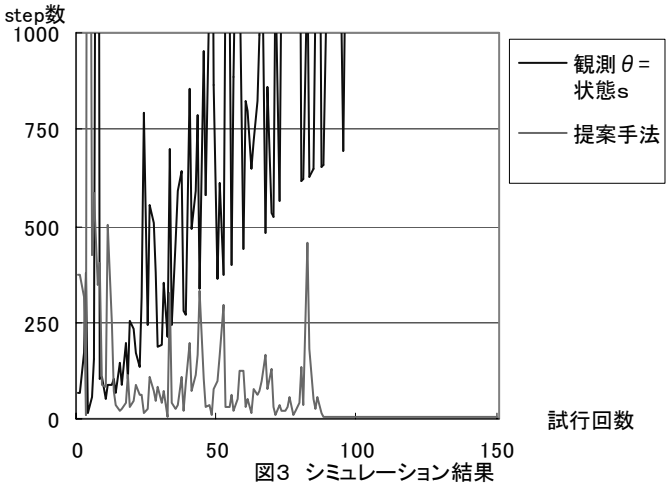

図3 シミュレーション結果

6) おわりに)

本研究では,) 部分観測環境下で,) 過去の系列から構成された状態より MDP を仮定した学習が可能であることを示した.) 今後はより現実環境に適応可能にするため,) 状態数の抑制手法,) 系列長の選択ルールの考案や実機への応用を予定である.)

参考文献)

[1] 土井幹也,小野功,小野典彦,木村元,小林重信：実環境への強化学習の適用に関する実験的考察,第 9 回インテリジェントシステムシンポジウム(1999).)

[2] 上原一寿,横井浩史,嘉数侑昇:任意の形態を有する機械のための学習制御システムの構築に関する基礎研究,) 情報処理北海道シンポジウム 2001(2001).)

[3] 井上康介,太田順,新井民夫:部分観測環境下での自律的状態空間構成を伴う実移動ロボットのナビゲーション行動(2002).)

[4] Richard) S.Sutton) and) Andrew) G.Barto,*Reinforcement) Learning:An) Introduction,*The) MIT) Press(1998):三上貞芳・皆川雅章共訳,強化学習,森北出版(2000).)

[5] Gail) A.Carpenter) and) Stephen) Grossberg,) ART2:Self-organisation) of) stable) cateGory) recognition) codes) for) analog) input) Patterns.Applied) Optics,vol.26,pp.4919-4930(1987).)