

知覚パターンの変化に基づく逐次学習型行動生成モデル

Sequential Learning Action Generation Model Based on Change in Sensory Pattern

郷古 学 伊藤宏司

Manabu Gouko Koji Ito

東京工業大学 大学院総合理工学研究科

Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

In this paper, we propose a sequential learning model that can generate behaviors for successfully performing various tasks. The model generates the actions based on change in sensory pattern. The model updates the memorized relationships between changes in sensory information and a motor command through sequential learning. We confirmed the performance of the model by applying it to a mobile robot. The results indicate that suitable behaviors for all the tasks emerged by the sequential learning.

1. はじめに

生物は実環境下において、さまざまな目的を達成することができる。これまでに、生物の持つこのような目標達成能力を有するロボットの実現を目指し、多くの研究が行われてきた。しかし、提案されるロボットのほとんどは、限定環境下で、指定されたタスク(目的)を実現するにとどまっている。

予測不可能に変化する実環境下で、さまざまなタスクを達成するためには、事前に、タスク達成に必要な行動系列(合目的的行動)を決定し、それを実現する行動生成モデル(コントローラ)を設計するのではなく、その時々々の環境・タスクに応じた行動を逐次的に生成し続けるメカニズムを実現しなくてはならない。

これまでに、状態-行動マッピングに基づく行動生成モデルが数多く提案されてきた。これらのモデルは、観測可能な変数(外部環境やロボットの身体の情報)により状態(知覚パターン) s を定義し、その知覚パターンがタスク達成時の状態(目標パターン) s^{goal} へと、直接または間接的に遷移するように、各状態に対し行動指令(モータコマンド) m をマッピングすることで、合目的的行動生成を実現するというものである。

設計時に想定した環境やタスクが変化するという事は、新規な知覚パターンや目標パターンが観測されることを意味している。行動生成モデルには、このような変化に対しても頑健なタスク達成が能力が必要とされる。

BrooksはSubsumption Architecture(以下、SAと呼ぶ)と呼ばれる、反射的な行動生成モジュール群の組み合わせにより、合目的的行動生成を実現するモデルを提案した[Arkin 98]。一般に、SA型のモデルは、環境変化にともない、新規な知覚パターンが観測されるような状況にもロバストに対応可能と言われており、学習により自動的にモデルを獲得する方法も提案されている[Maes 90]。しかし、そもそもSA型のモデルはタスクに依存して設計・学習を行う必要があるため、設計・学習時に想定していない新規タスクを実現することはできない。

これまでに筆者らも、行動生成モデルを提案し、その能力を検証してきた[Gouko 08]。このモデルでは、状態-行動マッピングではなく、行動と、それにより生じる知覚パターンの微

小時間の変化量のマッピングに基づき行動を生成する。そのため、環境が変化することで新規な知覚パターンが観測される場合であっても、知覚パターンの微小時間の変化量が同じであれば、行動生成には影響しないという特徴を持つ。

また、マッピングは自己組織的学習により、自動的に獲得(記憶)される。さらに、モデルが扱うタスクを学習時にあらかじめ設定する必要がなく、学習後に与えられる複数のタスクを実現することが可能である。これまでに、筆者らはモデルを移動ロボットへと適用し、学習後の同一モデルが、与えられた異なる2つのタスクを実現可能であることを示した。

しかし、このモデルでは学習により記憶したマッピングによって、ロボットのパフォーマンス(タスク達成の可否)が大きく変化するという問題を抱えている。この問題に対して、事前に網羅的にさまざまなマッピングを記憶しておく方法が考えられるが、学習時間や記憶容量の観点から現実的ではない。

そこで本稿では、これまでに筆者らが提案したモデル(以下、従来モデル)を基礎とし、新たに逐次学習機能を追加した行動生成モデルを提案する。提案モデルは、記憶しているマッピングを環境・タスクに合わせて逐次的に更新しながら行動を生成し続ける。これにより、限られた記憶容量であっても、安定したタスク達成が期待できる。本研究では、移動ロボットを用いた実験により、提案する逐次学習型モデルの能力を検証する。

本稿の構成は次のとおりである。次章では、従来モデルの動作メカニズムについて概説し、その後、逐次学習型モデルについて述べる。続いて、3章では、自律移動ロボットを用いた実験により、新たに提案した逐次学習型モデルの動作・性能を検証し、考察を行う。最後に4章でまとめを述べる。

2. 従来モデルと逐次学習型モデル

本章では、まず、筆者らが既に提案した知覚パターンの変化に基づく行動生成モデル[Gouko 08]の動作メカニズムを概説する。続いて、逐次学習型モデルについて説明する。

2.1 知覚パターンの変化に基づく行動生成モデル

モデルは、センサから得られる知覚パターンベクトル s をもとに、行動(モータコマンド)ベクトル m を出力する。各ベクトルを、それぞれ $s = (s_1, s_2, \dots, s_I)$ および $m = (m_1, m_2, \dots, m_J)$ とする。

行動主体(ロボット)が、時刻 t から $t + \delta t$ まで、モータコ

連絡先: 郷古 学, 東京工業大学 大学院総合理工学研究科,
〒226-8502 横浜市緑区長津田町 4259, 045-924-5654,
gouko@dis.titech.ac.jp

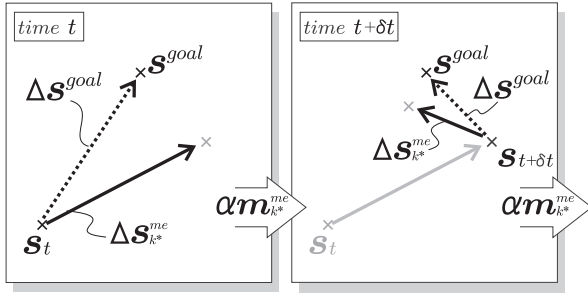


図 1: Mechanism of emergent behaviors.

マンド m により行動したとする (この間のモータコマンドの変化はないとする). 微小時間 δt における知覚パターンの変化 Δs は, 次式で表すことができる.

$$\Delta s = s_{t+\delta t} - s_t \quad (1)$$

この Δs は, 状態空間における, 行動 m による知覚パターンの微小時間の遷移方向を意味している. モデルは, 行動主体がさまざまな行動をすることで得られる, $(\Delta s, m)$ 対 (マッピング) を, あらかじめ学習により記憶しておき, それらを利用して合目的行動を生成する.

以下に, 学習により K 個の $(\Delta s, m)$ 対を記憶したモデルにおける行動生成手順を述べる. なお, モデルが記憶している $(\Delta s, m)$ のうち k 番目のものを $(\Delta s_k^{me}, m_k^{me})$ と表す.

まず, モデルに対して目標パターンベクトル $s^{goal} = (s_1^{goal}, s_2^{goal}, \dots, s_I^{goal})$ を与える. この目標パターンは, モデルがタスク達成時に観測される知覚パターンを意味している. モデルは, 自らの観測知覚パターンを目標パターンへと遷移させるように行動を生成する. s^{goal} と現在の時刻 t における知覚パターン s_t との差を Δs^{goal} とする (図 1 左).

次に, 記憶している $(\Delta s_k^{me}, m_k^{me})$ の中で, Δs^{goal} と最も類似している (ベクトル Δs^{goal} とのなす角が最小となる) ベクトル Δs_k^{me} を決定する. ここで, Δs_k^{me} と対になっている行動ベクトル m_k^{me} は, モデルが記憶している行動 m_k^{me} の中で s_t を, 最も s^{goal} へと近づけることが期待される行動を意味する. 時刻 t におけるモデルの実際の出力 m_t を,

$$m_t = (|\Delta s^{goal}| / |\Delta s_k^{me}|) m_k^{me} \quad (2)$$

とする. ここで $|\Delta s^{goal}| / |\Delta s_k^{me}|$ はゲインに相当する. Δs^{goal} が 0 ベクトルの場合には, $m_t = 0$ とする.

モデルは, この出力を微小時間 δt 持続した後, 時刻 $t + \delta t$ においても同様に, Δs^{goal} と最も類似しているベクトル Δs_k^{me} を求め, 行動を生成する (図 1 右). このように微小時間ごとに行動を生成することで, 行動主体の知覚パターン s が目標パターン s^{goal} へ近づく. つまり, タスクの達成が期待できる.

このモデルは, 学習時には特定のタスクの設定が不要であり, 目標パターンは行動主体が観測可能な任意のパターンとすることができる. そのため, 多様なタスクに対応することが可能である. これまでに筆者らは, 移動ロボットを用いた計算機実験により, 学習済みの同一モデルに異なる s^{goal} を与えることで, それぞれの目標パターンに対応して障害物回避や壁沿い直進という異なるタスクの実現が可能であることを示した [Gouko 08].

しかし, 一般に, ある行動により観測される知覚パターンの変化量は, モデルの置かれた環境によりさまざまである. そ

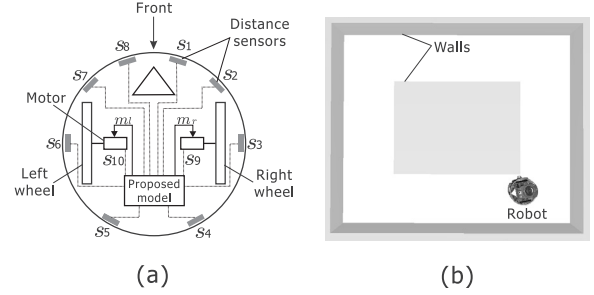


図 2: (a) Mobile robot. (b) Course.

のため, 記憶している $(\Delta s_k^{me}, m_k^{me})$ に基づいて行動を生成したとしても, 意図した (目標パターンへと近づく) 状態遷移が実現できず, 結果としてタスクが達成できない場合が生じる. この問題に対し, さまざまな環境下における $(\Delta s, m)$ を網羅的に記憶しておくことも考えられるが, 学習時間・記憶容量の点から現実的ではない. そこで, 行動生成時に観測される $(\Delta s, m)$ をもとに, 記憶している $(\Delta s_k^{me}, m_k^{me})$ を逐次的に更新する, 逐次学習機能をモデルに取り入れる. 逐次学習により, モデルは, 自らが置かれた環境下で観測される $(\Delta s, m)$ により, 自身の記憶を常に更新し, 行動生成に反映させることができる.

2.2 逐次学習型モデル

前述の従来モデルを基礎とした逐次学習型モデルについて説明する. 以下に, 逐次学習手順を述べる. まず, 時刻 t において, 式 (2) により出力 m_t を求め, δt 時間行動する. 次に, 時刻 $t + \delta t$ において観測される $s_{t+\delta t}$ と, 時刻 t において既に観測した s_t から $\Delta s = s_{t+\delta t} - s_t$ を求める. そして, 時刻 t で行動生成に使用した $(\Delta s_k^{me}, m_k^{me})$ を以下の式で更新する.

$$\begin{aligned} \Delta s_k^{me} &\leftarrow \Delta s_k^{me} - \eta(\Delta s_k^{me} - \Delta s) \\ m_k^{me} &\leftarrow m_k^{me} - \eta(m_k^{me} - m_t) \end{aligned} \quad (3)$$

ここで, η はある定数である. このように, 自らの行動とそれにより得られる知覚パターンの変化量を毎時更新し続けることで, 記憶している $(\Delta s_k^{me}, m_k^{me})$ が, モデルが現在置かれている環境下で, 実際に行動して得られた知覚パターンの変化量と異なる場合であっても, 逐次修正され, 行動が生成される.

3. 実験結果及び考察

逐次学習型モデルを移動ロボットシミュレーション及び実機へと適用し, 能力を検証した. 実験に用いた移動ロボットを, 上から見た図を図 2(a) に, シミュレーション環境を同図 (b) に示す. ロボットは直径 70mm, 高さ 40mm の円筒形で, 周囲に 8 つの距離センサを持っている. 各センサのレンジは 40mm で, 物体との距離 0 ~ 40mm を 0 ~ 1 の間で正規化した値をセンサ出力値とした. また, ロボットは独立に回転する 2 つの車輪を持っている.

実験において, モーターコマンド m は 2 次元 ($J = 2$) とし, 各要素は δt 時間ごとに左右のモータに独立に与えられる角速度 $(-108 \sim 108[\text{deg/sec}]$ を $-1 \sim 1$ の間で正規化した値) に対応する. 微小時間 δt は 0.5[sec] とした. また, 知覚パターン s は 10 次元ベクトル ($I = 10$) であり, $s_1 \sim s_8$ は 1~8 番のセンサ出力値に対応しており, s_9, s_{10} は左右のモータの角速度とした. ただし, 実験では簡単のため, $\Delta s, \Delta s^{goal}$ を求める際に, s_t の s_9, s_{10} の値を 0 とし, $s_{t+\delta t}$ 及び s^{goal} での s_9, s_{10} の値 (角速度) をそのまま用いた.

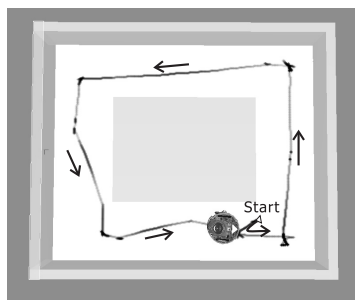


図 3: Trajectory of the robot (collision-avoidance).

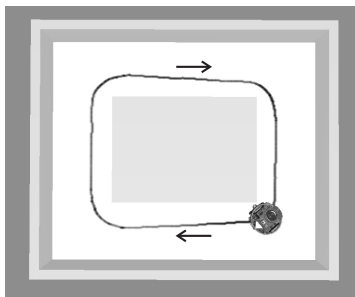


図 4: Trajectory of the robot (wall-following).

本実験において、モデルが記憶可能な $(\Delta s_k^{me}, m_k^{me})$ は全部で 4 つとした ($k = 1 \sim 4$)。各 m_k^{me} の初期値は、それぞれ $m_1^{me} = (1, 1)$, $m_2^{me} = (-1, -1)$, $m_3^{me} = (1, -1)$, $m_4^{me} = (-1, 1)$ とした。これらは順に、前進、後退、右回転、左回転に対応する。各 Δs_k^{me} の初期値はランダムとし、 $\eta = 0.7$ とした。

実験では、モデルに対して 2 つの異なる目標パターン s^{taskA} , s^{taskB} を与え、各目標パターンごとに期待されるタスクが実現可能かどうかを確認した。 s^{taskA} は s^{goal} の各要素のうち、 $s_9^{goal} = s_{10}^{goal} = 0.1$ とし、それ以外の要素を全て 0 としたパターンであり、すべてのセンサ出力値を 0 とし、かつ両車輪を同じ速度で前方へ回転させることを意味している。つまり、障害物回避タスク (collision-avoidance) を表わしている。

s^{taskB} は s^{goal} の各要素のうち、 $s_9^{goal} = s_{10}^{goal} = 0.1$, $s_3^{goal} = 0.4$ とし、それ以外の要素を全て 0 としたパターンである。 s^{taskB} は、3 番のセンサ出力値を一定に保ちながら直進すること意味しており、これは進行方向の右側の物体との距離を保ったまま直進、つまり、壁沿い直進タスク (wall-following) を表わしている。

目標パターンとして s^{taskA} , s^{taskB} を与えた場合の、ロボットの軌道の一部を図 3, 4 に示す。図 3 を見ると、ロボットの周囲に壁が存在しない場合は、直進軌道を示しており、壁に接近すると、方向転換を行い回避している。

また、図 4 を見ると、ロボットが壁沿いに前進している様子が確認できる。いずれの実験においても、初期値として設定したランダムなマッピングでは、タスクを達成できないことを確認しており、逐次学習の効果により、タスクが達成できたと言える。

同様の実験を実機を用いて行った様子を図 5 に示す。これは、障害物回避タスク実行時のロボットの 1sec 毎のスナップショットである。方向転換により、壁を回避している様子が確

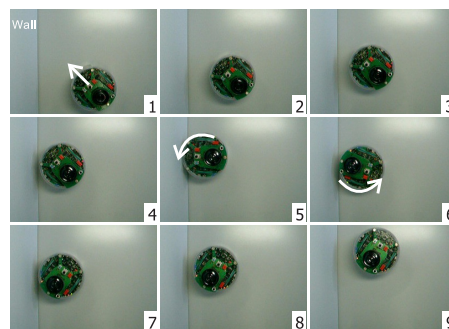


図 5: Snapshots of the collision avoidance behavior generated by the real robot.

認できる。

その他の実験結果として、壁沿い前進タスクにおいて、タスクを行っていたものの、途中から壁から離れていってしまう等のロボットの挙動を確認した。ロボットの挙動は式 (3) で示す更新式の定数 η により大きく変化する。この定数が大きいと、更新量が大きく即応性が期待できるが、挙動が不安定になる傾向がある。逆に、 η が小さいと、逐次学習による更新量が小さいため、モデルが現在置かれている環境の情報を反映した行動生成が困難になる。今後は、 η を動的に変化させる、新しい逐次学習法を検討する必要がある。

4. まとめ

本稿では、知覚パターンの変化に基づく行動生成モデルに逐次学習機能を追加した、新たな行動生成モデルを提案し、移動ロボットを用いた実験によりその能力を検証した。提案モデルは、記憶しているマッピングを、モデルが現在置かれている環境下で得られる情報に基づき、逐次的に更新することで、常に環境の情報を反映した行動生成が可能となり、頑健なタスク達成能力が期待できる。実験の結果、初期値としてランダムなマッピングを設定したにもかかわらず、タスク達成が可能であることを確認した。今後の課題として、逐次学習則の改良を予定している。

謝辞

本研究の一部は人工知能研究振興財団及び柏森情報科学振興財団の助成によるものである。ここに謝意を表す。

参考文献

- [Arkin 98] Arkin, R. C.: *Behavior-Based Robotics*, The MIT Press (1998)
- [Gouko 08] Gouko, M., Tomi, N., Nagano, T., and Ito, K.: Behavior emergence model for performing multiple tasks, *Proceedings of 2008 IEEE International Conference on Distributed Human-Machine Systems (DHMS2008)*, pp. 491-496 (2008)
- [Maes 90] Maes, P. and Brooks, R. A.: Learning to Coordinate Behaviors, *Proceedings of the Eighth National Conference on Artificial Intelligence (AAAI'90)*, pp. 796-802 (1990)