

ロボットの身体バブリングによる身体制御則の獲得のモデル化

Modeling of Self Body Control Acquisition using Motor Babbling by a Robot

青木 達哉

Tatsuya Aoki

西原 成

Joe Nishihara

中村 友昭

Tomoaki Nakamura

長井 隆行

Takayuki Nagai

電気通信大学

The University of Electro-Communication

Human intelligence is deeply dependent on its physical embodiment, and its development requires interaction between its own body and surrounding environment. However, it is still an open problem that how we can integrate the lower level motor control and a higher level symbol manipulation system. One of our research goals is to make a computational model of human intelligence from the motor control to the higher level symbol manipulation. To this end, we propose a robot motor control learning as the first step in this paper. The method is based on iHMMs (infinite Hidden Markov Models). The robot moves its arm randomly by changing torques of joint angles and obtains the pose of its arm. The iHMM uses state space for representing the relationship between joint torques and pose of the arm by segmenting the obtained sensory-motor information autonomously. The robot can gradually learn to move its arm to a specific position by planning the torque sequence using the learned model.

1. はじめに

近年、人工知能の発展が目覚ましく、特に深層学習のパターン認識タスクにおける成功には目を見張るものがある。例えば、畳み込みニューラルネットワーク (CNN) [1] は、特定のタスクにおいて人の視覚に迫る、もしくは既に超えているケースすら存在する。また、深層強化学習 (DRL, DQN) は、様々なビデオゲームの戦略を試行錯誤を通して学習し、人のハイスコアを超えることが示されている [2, 3]。

一方で、人間の知能は身体と不可分であり、その発達には身体と環境の相互作用が不可欠であると考えられる。Levine らは、こうした身体と環境の相互作用を基盤とした知能を感覚運動系の強化学習という形で追究しており、例えばロボットが子どものおもちゃを使って遊ぶような意味のある行為を学習できることを示している [4]。しかしながら、依然として身体的な制御から言語操作の様な高次の知能までを結びつける統合的な仕組みは実現されておらず、難しい問題であると言える [5]。

本研究は、このような身体から記号までを統合するモデルを構築し、ロボットの知的な行動や言語理解を実現することが究極の目標である。本稿では、そのための要素として、特に身体と環境の相互作用を基盤とした真の知能を実現するための身体制御則の獲得について検討する。重要なポイントは感覚運動学習であるが、これを幼児が試行錯誤的かつインクリメンタルに学習するように、モーターバブリングによって獲得するアルゴリズムを構築することを目指したい。本稿では、隠れマルコフモデルによってこうした学習を実現することを検討する。提案する隠れマルコフモデルは、ロボットがランダムに動かし腕の軌道やその際の各関節のトルクを分節化し、状態空間で表現する。そして現在の手先の位置から、目標座標に腕を動かすための最尤な状態系列を推定することで、腕を制御するためのトルク時系列を計算することができる。

このような確率モデルによる身体制御のモデル化は、筆者らのグループが進めている言語獲得過程の確率モデル [6] との確率的な統合が容易であり、身体から言語まで全体を統合するために有効であると考えている。さらには、[7, 8] で提案され

ている強化学習への展開を取り入れることで、試行錯誤によって身体的な制御を獲得し、そこから徐々に高次の記号や言語操作にまで至る知能を実現できる可能性があると考えている。本稿では、モーターバブリングによる身体制御則の獲得を基本的な要素として、多層マルチモーダル LDA (mMLDA) や強化学習を統合した知能システム全体の構成についても議論する。

モーターバブリングからのロボットによる感覚運動学習に関しては、従来いくつかの研究が存在する [9, 10, 11]。しかし、これらを記号操作にまで接続しようとしている試みはみられない。一方、高野らは、身体運動から言語までを統一的に扱うモデルを提案している [12]。しかしロボットの様な主体が、身体バブリングによって制御則を獲得し、試行錯誤から徐々に言語使用のような高次の知能につなげるインクリメンタルなモデルは考えていない。

2. 隠れマルコフモデルによる動作のモデル化

ここでは、まず、モーターバブリングによって取得されるデータを構造化することで、身体制御を学習する手法を検討する。状態に対して、関節トルクを与えると、状態に応じて、関節角度が観測情報として、得られるモデルを考える。このモデルでは、マルコフ性を仮定し、どのようなトルク値を与えれば、目的の姿勢になるかを表現している。

2.1 状態空間とトルク指令値の学習

提案手法の学習は、(1) 状態空間の学習と、(2) 各状態におけるトルク指令値の学習の 2 段階に分けて行う。状態空間の学習では、関節角を観測情報とする infinite Hidden Markov Model (iHMM) [13] を用いることで、状態空間を構築する。この学習によって、ロボットが取りうる連続的に変化する姿勢を、離散化した隠れ状態で表現することができる。

続いて、トルク指令値の学習を行う。(1) の学習を行うことで、観測データがどの状態であるかは推定済みである。この観測データにおいて、状態遷移が起きた際に与えたトルク値に注目すると、ある状態 s のとき、別の状態 s' に遷移させる関節トルクに関する観測情報を抽出できる。このデータを用いて、各状態間の遷移に必要なトルク値に関する確率分布を推定する。ここでは、この確率分布はガウス分布で表現できると仮定

連絡先: 青木 達哉, 電気通信大学大学院情報理工学研究所,
東京都調布市調布ヶ丘 1-5-1, aoki@apple.ee.uec.ac.jp

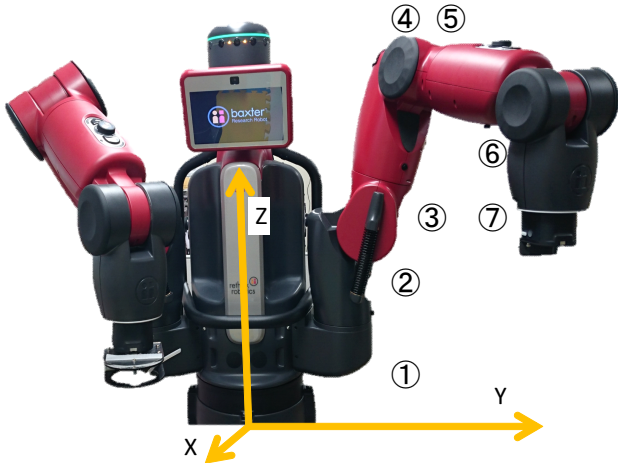


図 1: Baxter

する。

2.2 状態系列の推定

手先を目標点まで動作させる状態系列の推定は、以下の式 (1) で求められる確率が最大になる状態系列を、動的計画法の 1 つであるビタビ・アルゴリズムを用いて求めることを行う。

$$P(s_0, s_1, \dots, s_N | \mathbf{X}_s, \mathbf{X}_g) \propto P(s_0) P(\mathbf{X}_s | s_0) P(\mathbf{X}_g | s_N) \prod_{t=0}^{N-1} P(s_{t+1} | s_t) \quad (1)$$

なお、 \mathbf{X}_s は手先の動作開始点の座標、 \mathbf{X}_g は手先の目標点の座標、 N は状態遷移回数をそれぞれ表す。ただし、ここでは簡単のため、状態遷移を考える際には自己遷移が起きないという制約を加え、状態遷移回数 N は事前に決定した値とする。実際には、モデルの学習に速度の情報を含めることで、こうした制約を回避できる。

3. 実験

rethink robotics 社製の双腕作業ロボット Baxter のシミュレータを用いて、提案手法による制御則の学習と経路生成の実験を行った。図 1 に示すように、Baxter の腕は 7 自由度を持つ。

学習データを収集する際は、ロボットの腕の各関節に作用させるトルク値をランダムに決定し、一定時間その値で腕を動作させた。この試行の中で、ロボットに作用させた関節トルクと関節トルクを作用させた結果、観測される関節の角度 \mathbf{X} を記録した。この試行を、1000 回繰り返し動作させて学習データの取得を行った。図 2 が、ロボットが実際に動作した際に観測された関節角に対応する、手先の x, y, z 座標をプロットしたものである。上述のように、観測された関節角で状態空間を学習すると、図 3 のような状態の中心で表現される、離散的な 42 個の状態に分けられた空間が学習された。

続いて、学習された状態空間で、2.2 節で述べた方法により目標点への腕の経路計画の実験を行った。提案手法で計画した経路はランダムに動作した結果で獲得した状態の遷移をもとに生成されているため、必ずしも目標へ最短で繋がる姿勢変更の経路を推定できるわけではない。しかし、状態系列を求めることで目標姿勢に向かう経路計画が行えていることを確認した。

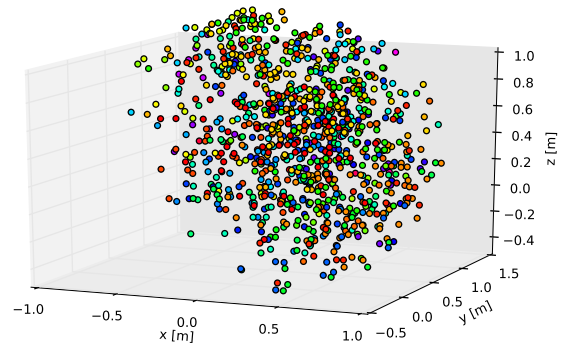


図 2: ロボットが経験した動作点

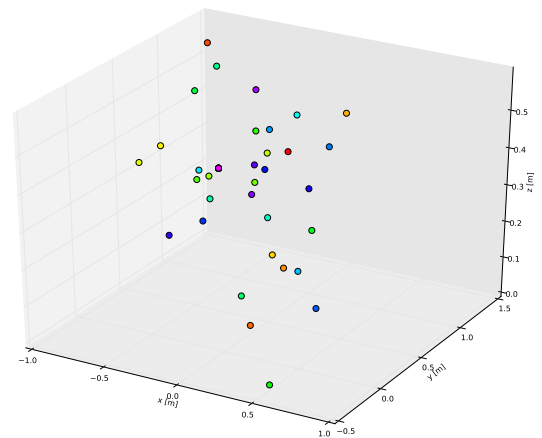


図 3: ロボットが学習した状態空間

さらに、学習時にある状態から異なる状態へ遷移する際に各ガウス分布の平均をロボットの動作の命令値とし、トルク制御でロボットの腕を動かすことが可能である。選択されたトルク値でのロボットの腕の軌道の評価や滑らかな軌道の生成が今後の課題になる。

滑らかな軌道を計画するためには、状態間の遷移の多様性が重要となる。つまり、空間的に隣り合う状態の遷移の多様性を増やすために、腕を動かす経験の多様性が必要である。そうした意味では、ランダムな身体バブリングを多く実行することは有効であると考えられる。そして、こうしてできあがった状態空間に基づいて、実際の環境でタスクを実行することで、より環境やタスクにとって有効な動きが計画できるような空間に更新していくことが、次の学習ステップであると考えている。

4. 身体制御から言語学習までの統合

ここでは、提案した身体制御モデルを基盤として、身体バブリング、強化学習、概念学習、言語学習までを統合するモデルについて考える。

4.1 多層 LDA

筆者らは、多様な概念を形成するために、異なる種類の入力を範疇化する複数の MLDA を下位層とし、上位層にはそれらを統合する MLDA を配置することで概念間の関係を表現する多層 MLDA (multilayered MLDA: mMLDA) を提案した [6]。図 4 に、mMLDA のグラフィカルモデルを示す。この

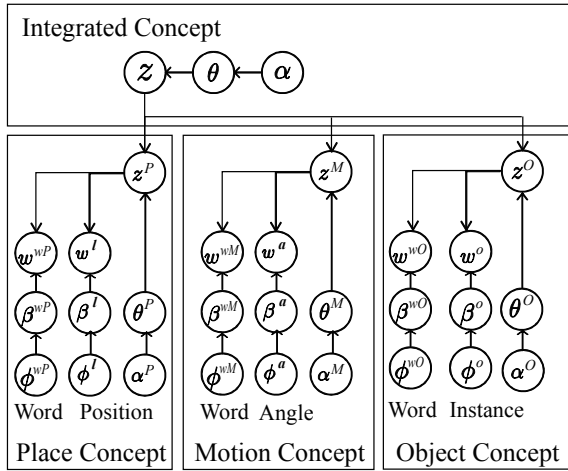


図 4: mMLDA のグラフィカルモデル。

モデルは、下位層に物体、人の動き、場所の概念を形成する MLDA を、上位層にそれらを統合する MLDA を配置している。ただしこれは一例であり、下位層にどのような概念が形成されるかは、観測する情報に依存する。例えば、人を検出してその特徴を観測すれば、男性、女性、子どもや大人といった、人の見た目に関する概念が形成される [6]。

図 4 において、 z は統合概念を表すカテゴリであり、 z^P , z^M , z^O はそれぞれ下位概念 (概念クラス) に相当する、物体、動き、場所カテゴリである。上位カテゴリ z は、下位概念間の関係、つまりは共起性を表現したモデルとなっている。また、 w^o , w^a , w^l は、それぞれ物体に関する情報、物体を扱っている際の人の動きの情報 (関節角など)、位置情報である。さらに、 w^{wO} , w^{wM} , w^{wP} は、教示発話から得られる単語情報である。 ϕ^* , α^* は、ハイパーパラメータである。

文献 [6] の例では、観測情報はウェアラブルセンサを用いて取得している。しかし、ロボットの言語獲得を考えた場合は、ロボット自身の関節角や視聴覚・触覚情報などが入力になる。

さらに文法を獲得するためには、統語範疇と統語規則を学習する仕組みが必要である。ここでは最も簡単な考え方として、統語範疇を概念の種類 (概念クラス) と対応させ、文章における統語範疇の順番を統語規則と定義する。筆者らは、mMLDA とベイジアン隠れマルコフモデル (Bayesian Hidden Markov Model: BHMM) を組み合わせることでこれらを学習する仕組みを提案した [6]。ここでのポイントは、マルチモーダルで多様な概念と統語範疇、および統語規則の相互依存的な学習である。まず mMLDA を学習し、相互情報量によって各単語の統語範疇 (概念クラス) を決定する。この統語範疇を初期値にして、教示文を使って BHMM を学習する。従って、BHMM によって統語規則と統語範疇が学習されることになる。統語範疇が更新されるため、その情報を用いて再度 mMLDA を学習する。これを繰り返すことにより、文法と概念を相互に精緻化することができる。

4.2 強化学習への展開

図 4 のグラフィカルモデルを見ると、ロボットの行動を表現する上位層と、下位の各概念の組み合わせで表現される隠れ状態、及び観測データによって mMLDA が構成されていることが分かる。従って、これを時間軸方向に接続することで、部分観測マルコフ決定過程 (POMDP) と構造の似たモデルを構築することができる。図 5 にこの様子を示す。図の左側は、

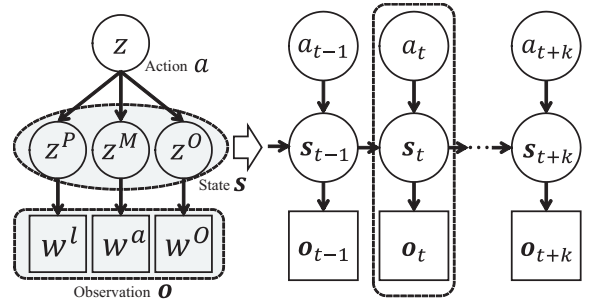


図 5: mMLDA の時間発展: 左) mMLDA のグラフィカルモデル (変数の定義は 4 に同じであるが、簡単のために分布やハイパーパラメータは省略している)。右) mMLDA を時間発展させたモデル。

mMLDA の分布やハイパーパラメータを省略し、行動、状態、観測で表現したものであり、右側はこれを時間的に接続したものである。この段階で行動 a_t は隠れ変数であり、観測できないことに注意が必要である。つまり、ある状態においてどのような行動を取ったのかは確率的な予測であり、観測 o_t から推論する必要がある。

状態 s_t も同様に隠れ変数であり、各下位概念の組み合わせになっている。このように状態を下位概念の全ての組み合わせと考えることで、このモデルの学習は一般的な HMM の学習 (EM アルゴリズム) として定式化できる。

概念学習は、強化学習と同時に進められる時空間分節化・範疇化であると言える。ここでは簡単のために、強化学習によって蓄積されたマルチモーダルデータを使ってバッチ学習することを考える。まず mMLDA の学習は、データ全体を mMLDA によって段階的にカテゴリ分類することで実現される。この後に、HMM の学習を行う。mMLDA によってある種記号化された時系列データを、時間的に分節化しつつカテゴリ分類するのが、HMM のパラメータ推定である。

次に、パラメータ推定後のモデルを用いた行動計画について考える。入力される信号は、現在時刻 $t=0$ の観測 o_0 であり、最終的な目的状態へ遷移するための最短の行動系列 $\mathbf{a} = \{a_0, a_1, \dots, a_n\}$ を求めることが目標となる。この問題は、時間 t 、状態 s_t 、行動 a_t を軸としたトレリス空間で、最尤となるビタビパスを求める問題と考えることができる。ただし、観測データは実際に行動を取らない限り得られないため、現在の観測データとその後は、予測される観測データを用いる。トレリス上で目標となる状態にたどりついたとしても、その確率が、同じ時間における他の状態に存在する確率に比べて小さい場合には、目標の状態にたどりついていない可能性が高い。そこで、目標の状態に他の状態に比べて十分に高い確率で存在する場合に、行動計画が終了したとしてビタビパスをバックトラックする。

以上の手法によって、行動系列 \mathbf{a} を計画することができる。しかし、実際には行動 a_t は概念としての行動であり、実際にロボットが行動するためには、具体的な動作を決定する必要がある。決定すべき具体的な動作は、図 5 における w^a に相当する。これは本来可観測であるが、将来の行動計画においては未観測であり、観測情報より推定する必要がある。つまり、 $\hat{w}_t^a = \arg\max_{w^a} p(w^a | z_t, z_t^P, z_t^M, z_t^O)$ を解けばよく、これは mMLDA の枠組みで計算可能である。また、他の観測データについても同様に予測することができる。実際に行動を行った後に、予測した観測データ・状態と実際の観測データ・状態とのずれを計算することができるため、このずれが大きい場合

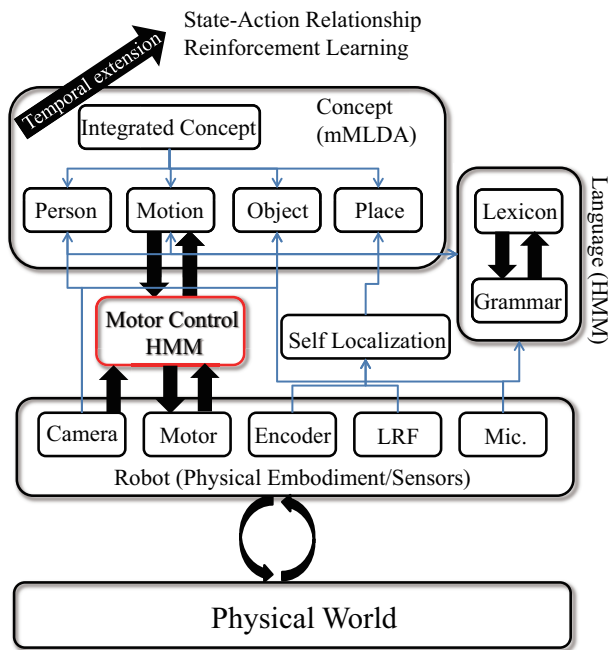


図 6: 全体を統合するモデルの概念図.

にはプランニングを行い、新たな計画に従って動作を決定する。

4.3 全体の統合

最後に、以上の要素を統合して、身体から言語までを統合することを考える。図 6 に全体を統合したモデルの概念図を示す。ロボットは、養育者を含む物理的な世界と相互作用しながら、徐々に身体の制御を獲得する。同時に、概念学習により、行動や物体、人などの概念や語彙、文法を学習する。さらには、強化学習により概念の組み合わせで表現される状態空間と、行動の対応関係を学習する。重要なことは、これらが相互につながっており、お互いに作用を及ぼしあいながら学習が進んで行くことである。また、モデルのパラメータは、バッチ学習ではなく経験と共に得られるデータをオンラインで学習する。これによって、身体制御から言語使用までを発達の的に学習する知能をロボットを実現できるのではないかと考えている。

5. まとめ

本稿では、幼児が試行錯誤運動から身体制御則を学習する過程を隠れマルコフモデルによって実現できるかを検討するために、各関節のトルク値を入力、関節角度を出力とするモデルで状態空間の学習を行った。状態空間内で適切な状態に遷移することで、目標位置に近い状態に遷移するような経路が生成できることが確認できた。即ち、身体制御則を学習する過程を隠れマルコフモデルによってモデル化できる可能性があると言える。実際にロボットにこのアルゴリズムを搭載し、生成された経路通りに遷移させられるかについて検証を続ける。

また本稿では、身体制御から言語までを統合するモデルについて検討を行った。今後、こちらについても実際にモデルをロボットに実装し、実験を行う予定である。

謝辞

本研究は、JST CREST の支援を受けたものである。

参考文献

- [1] A. Krizhevsky, S. Ilya, G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems 25, pp.1097–1105 (2012)
- [2] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller, "Playing Atari with Deep Reinforcement Learning," NIPS Deep Learning Workshop (2013)
- [3] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, "Human-level Control through Deep Reinforcement Learning," Nature 518, pp.529–533 (2015)
- [4] S. Levine, C. Finn, T. Darrell, P. Abbeel, "End-to-End Training of Deep Visuomotor Policies," arXiv 1504.00702 (2015)
- [5] T. Taniguchi, T. Nagai, T. Nakamura, N. Iwahashi, T. Ogata, H. Asoh, "Symbol Emergence in Robotics: A Survey," arXiv 1509.08973 (2015)
- [6] M. Attamimi, Y. Ando, T. Nakamura, T. Nagai, D. Mochihashi, I. Kobayashi, H. Asoh, "Learning Word Meanings and Grammar for Describing Everyday Activities in Smart Environments," EMNLP 2015, pp.2249–2254 (2015)
- [7] 長井, 中村, アツタミミ, 持橋, 小林, 麻生, "多層マルチモーダル LDA と強化学習による意味理解に基づく行動決定," 人工知能学会全国大会 2015, 2F4-OS-01a (2015)
- [8] 長井, 中村, "記号創発ロボティクス —マルチモーダルカテゴリゼーションから言語に至る構成の道筋—," 人工知能, Vol. 31 No. 1 (2016)
- [9] Y. Demiris, A. Dearden, "From motor babbling to hierarchical learning by imitation: a robot developmental pathway", Proceedings of the fifth international workshop on epigenetic robots (EPIROB): modeling cognitive development in robotic systems, pp. 31–37 (2005)
- [10] R. Saegusa, G. Metta, G. Sandini, S. Sakka, "Active Motor Babbling for Sensorimotor Learning," IEEE International Conference on Robotics and Biomimetics, pp. 794–799 (2008)
- [11] 山田, 高橋, 尾形, チャンドラ, 菅野, "身体バブリングによる事前学習を用いた柔軟関節ロボットの効果的な動作生成", 情報処理学会第 77 全国大会, pp.513–514 (2015)
- [12] W. Takano, Y. Nakamura, "Statistical Mutual Conversion between Whole Body Motion Primitives and Linguistic Sentences for Human Motions," The International Journal of Robotics Research, 34, pp.1314–1328 (2015)
- [13] M. J. Beal, Z. Ghahramani, C. E. Rasmussen: The infinite hidden markov model, Advances in Neural Information Processing Systems, vol. 14, pp. 577–584 (2002)