

多層マルチモーダルLDAと強化学習による 片付け概念獲得への取り組み

An Approach to Concept Acquisition of Tidying up using Multilayered Multimodal LDA and Reinforcement Learning

恒川英里 *1
Eri Tsunekawa

ムハンマド アッタミミ *4
Muhammad Attamimi

小林一郎 *1
Ichiro Kobayashi

麻生英樹 *2
Hideki Asoh

持橋大地 *3
Daichi Mochihashi

中村 友昭 *4
Tomoaki Nakamura

長井 隆行 *4
Takayuki Nagai

*1 お茶の水女子大学
Ochanomizu University

*2 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

*3 統計数理研究所
The Institute of Statistical Mathematics

*4 電気通信大学
The University of Electro-Communications

It is necessary for robots to be practically used in the real world, they have to be able to behave properly to the given task according to the information they observe with their sensors. In this study we take a basic study on the acquisition of robot's action knowledge to the given task by means of Q-learning based on the recognition of the visual information. In concrete terms, we use a humanoid robot HIRO made by Kawada Industries, Inc. and make it acquire the concept of tidying up. We have introduced Multilayered Multimodal Latent Dirichlet Allocation to acquire the concept of moving an object particular place and employed Partially Observable Markov Decision Process to make the sequential states observed by mMLDA to acquire the total concept of tidying up. We show the process of acquiring the concept of tidying up by conducting an experiment with a humanoid robot.

1. はじめに

ロボットと人が一緒に暮らす世界が実現しつつある中、ロボットが人との生活に適応するためにはロボットにも意思疎通の能力が必要不可欠である。また、人の仕事を代行するという役割を果たすことも今後期待される。本研究では、ロボットが前提知識がない状態から、人が行うような片付けという行動知識を獲得するモデルを実現することによって、ロボットの知能化に挑戦する。

ロボットの知能化について、人間が、行動の学習、計画する仕組みをモデルに提案された長井らの先行研究 [5] がある。このモデルは様々なレベルでのプランニングや行動決定、言語などを統一的に扱うことができる枠組みを構築することを最終的な目標としている。本研究ではこのモデルを基にマルチモーダル情報を用いて概念獲得を試みる。ロボットが学習をする際、行動決定においては、センサ情報に基づく即時的なものから記号のような抽象度が高く、汎用性の高い仕組みを使った中長期的なプランニングに基づいたものまで様々なものがある。それらは大別すると、未知のものに対するモデルフリーの行動決定と、モデルベースの行動決定と見なすことができる。また、それぞれの学習フェーズが交互に入れ替わるなどの2段階による学習が行われている。片付けという課題に対しても、見たことない物体に対してはそれぞれどのように掴むのか、そして積めるものなのか、立て掛けるものなのかなどのモデルフリーの行動決定と全体のバランスを見て最適な位置に配置することのできるようなモデルベースの行動決定による行動計画の2段階の学習が必要と考えられる。

本研究では上半身ヒューマノイドロボットを用いて机の上の物体を整理することによって物体ごとに片付け概念を獲得し、人の思考と同様に、概念を用いて行動を計画立案する課題に、多

層マルチモーダルLDA (mMLDA: Multilayered Multimodal Latent Dirichlet Allocation) [4] と強化学習を用いて取り組む。

2. 片付け概念獲得

2.1 概要

テーブルの上に散乱している物体を所定の位置に移動させるという「片付け」概念の獲得を行う。図1に概要図を示す。机の上に無作為に置かれた物体をヒューマノイドロボットHI-RONXCのハンドカメラによって画像認識をし、その結果に基づき、Q学習を用いて、どの物体がどこに置かれるべきなのか、また、置くための行動概念を学習する。そして、その行動に対して、mMLDAを用いて概念を学習させ、部分観測マルコフ決定過程 (POMDP: Partially Observable Markov Decision Process) の枠組みを用いて行動計画を立案する。

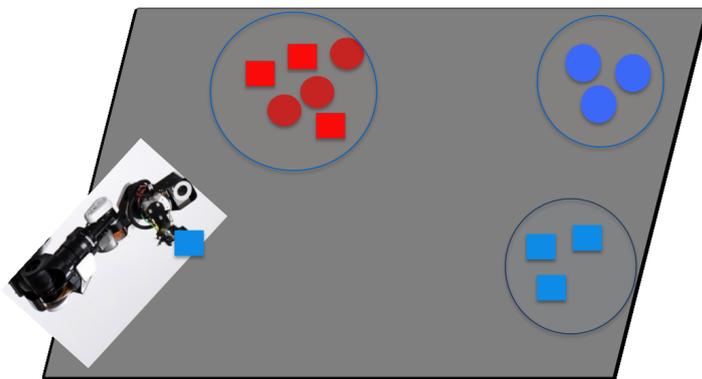


図 1: 片付け行動概要

連絡先: 恒川英里, お茶の水女子大学大学院人間文化創成科学研究科理学専攻情報科学コース小林研究室, 〒112-8610 東京都文京区大塚 2-1-1, tsunekawa.eri@is.ocha.ac.jp

2.2 mMLDA を用いた概念の取得

まず、物体ごとの正しい片付け知識を Q 学習を用いて学習する。次に、得られた物体に対する行動系列に、mMLDA を用いた言葉の概念の獲得を行う。今回、下位概念として使用するカテゴリは、物体、場所、行動の3つである。図2に mMLDA のグラフィカルモデルを示す。mMLDA とはマルチモーダル LDA (MLDA) [6] を多層化したモデルであり、このモデルの統合概念は教師なし学習によって学習することができる [4]。

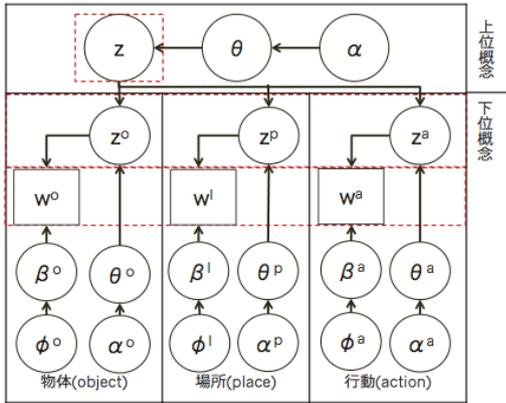


図 2: mMLDA のグラフィカルモデル

2.3 POMDP を用いた概念による行動計画

観測したマルチモーダルデータから、POMDP を用いて、観測情報から獲得した概念の選択、行動計画を行う。これは mMLDA における下位概念が POMDP における状態、上位概念が行動に対応していると考えられるからである。mMLDA と POMDP における関係を図 3, 4 に示す。mMLDA を時系列順に連結することによって POMDP モデルに置くことができることを示している。ここを学習することによって最終的な最適片付け行動のプランニングを行うことができる。この時 \bar{a} は概念 a の実際の行動を示している、行動知識の獲得に Q 学習を用いた POMDP の枠組みを図 5 に示す。

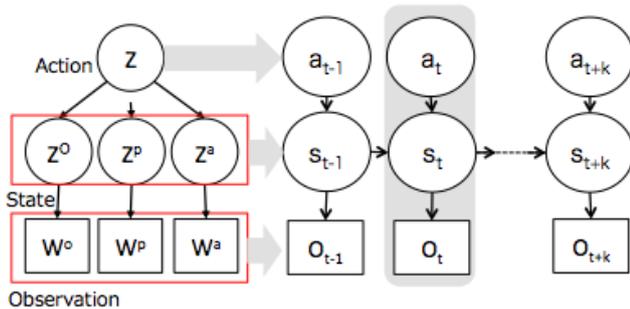


図 3: mMLDA の時系列連結による POMDP の構成

3. 実験

3.1 作業課題

図 1 に示したように Q 学習によって物体に対しての正しい場所を獲得し、その置くための行動概念を mMLDA で獲得したのちに正しい片付け行動系列の取得を行う。

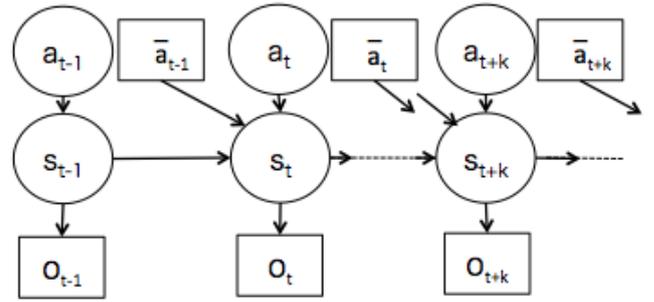


図 4: mMLDA を状態推定に用いた POMDP モデル

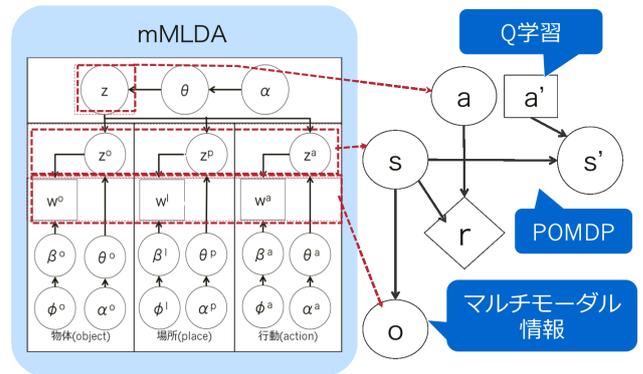


図 5: 行動知識獲得の方法

3.2 mMLDA を用いた行動概念の取得

mMLDA による行動概念獲得は対象とする行動概念を達成する動作の学習と他のモダリティ（この場合片付け対象となる物体がもつモダリティ）を組み合わせることで概念を獲得している。

3.2.1 画像処理による物体取得

ハンドカメラから得られた画像から OpenCV を用いて画像処理を行い、色認識、形状認識を行う（図 6, 7）。

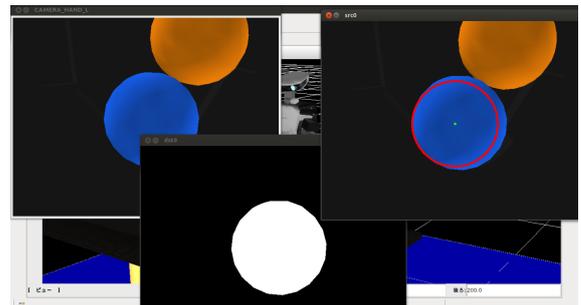


図 6: HIRO による画像処理（丸い物体の認識）

3.2.2 mMLDA における課題設定

机の上をグリッドとし、一つの物体に対して片付け位置を学習、それを複数物体に適用することによって片付け行動を行う。ロボットの行動を学習する Q 学習において、本来なら関節角の情報を用いて学習することが一般であるが、今回は課題

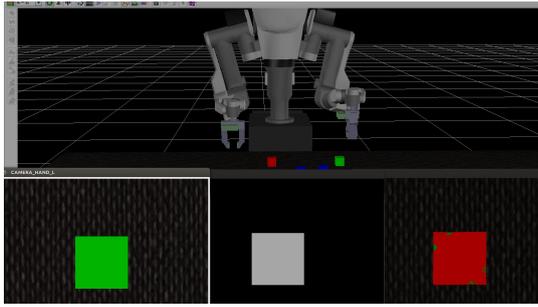


図 7: HIRO による画像処理 (四角い物体の認識)

を単純化して捉え、「掴む」、「置く」、の行動設定は組み込んだ形とした。3つに分類されるのマルチモーダル情報のうち、移動する物体の情報はロボットがハンドカメラで認識したものを対象とし、動作については上述のように「掴む」、「置く」が組み込まれた移動を対象として、適切な場所に物体を移動する Q 学習を行い、行動知識を獲得する。

状態、行動、報酬を以下のように設定する。

- 状態：円形度，rgb 値，物体の面積，場所（机の上の座標），動き
- 行動：(上，下，右，左)に移動する，
- 報酬：正しい場所に置かれたら正の報酬（10）

次に、今回、カテゴリ数は、物体、場所、行動の3つとあらかじめ決定し、物体を Q 学習課題の状態における、円形度，rgb 値，面積，場所を場所，行動を動きと対応させている。また、片付けるべき場所の名称はそれぞれ、机の上全体に対して「左上」、「右上」、「右下」、「その他」という名称で概念の獲得を行った。

下位概念の候補

- 物体：「赤く丸い」、「赤く四角い」、「青く丸い」、「青く四角い」、「なし」の5種類
- 場所：「左上」、「右上」、「右下」、「その他」、の4種類
- 行動：「掴む」、(左上，右上，右下，その他)に移動する、「置く」、「何もしない」、の7種類

上位概念の候補

- 「左上に赤い物体を置く」、「右上に青く丸い物体を置く」、「右下に青く四角い物体を置く」、「その他」、の4種類

3.2.3 Q 学習による概念獲得の結果

およそ 20 回の試行で収束していることがわかる。図 8 に Q 学習の結果を示す。

表 1 は 4 つの上位概念を獲得した際のそれぞれのモダリティの予測精度を表す。mMLDA を用いて行動の獲得を行った。獲得した概念は下位概念のカテゴリ 3 つと上位概念の 4 種類であり、上位概念について、4 つの候補のうち、一番スコアが高かったものをその行動概念とみなし、その識別精度は 95% となった。

3.3 獲得された概念を用いた行動計画

今回は課題の単純さから行動計画立案に POMDP ではなく頻度情報に基づくバイグラムを構築することにより、最適な行動系列を求めるとした。机の上にランダムに 5 個の物体が置かれている状態から片付ける行動を 1 エピソードとし、100 回

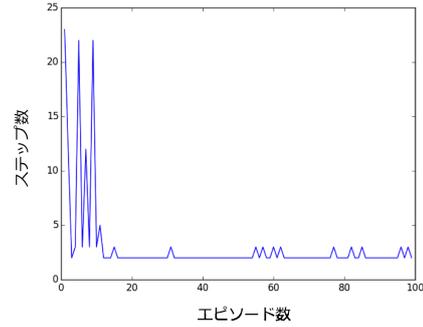


図 8: 学習結果

表 1: 概念獲得結果

カテゴリ	予測精度
物体	0.694
場所	1.000
行動	0.57
上位概念	0.95

行動系列を獲得したのち、バイグラムモデルを構築した。結果を表 2 に示す。今回、実験環境の都合上、「その他」という概念の選択は行なわれていない。紙面の都合上、概念の「物体を置く」部分は、省略している。

表 2: シミュレーション結果

統合概念	確率
< s > 右上に青く丸い	0.20
< s > 右下に青く四角い	0.20
< s > 左上に赤い	0.60
右上に青く丸い 右上に青く丸い	0.31
右上に青く丸い 右下に青く四角い	0.20
右上に青く丸い 左上に赤い	0.49
右下に青く四角い 右上に青く丸い	0.21
右下に青く四角い 右下に青く四角い	0.26
右下に青く四角い 左上に赤い	0.53
左上に赤い 右上に青く丸い	0.50
左上に赤い 右下に青く四角い	0.25
左上に赤い 左上に赤い	0.21
右上に青く丸い </s >	0.22
右下に青く四角い </s >	0.27
左上に赤い </s >	0.32

今回のシミュレーションは観測物体に対して、100%の確率で正しい統合概念を獲得し、行動できるとし、この結果を基にプランニングを行い、「左上に赤い物体を置く」、「右上に青く丸い物体を置く」、「右下に青く四角い物体を置く」、「左上に赤い物体を置く」、「左上に赤い物体を置く」という結果が得られた。

3.4 考察

今回、簡単にはあるが、観測物体に対しての片付け行動を Q 学習によって獲得、それに対する mMLDA を用いた統合概念の形成、概念によるプランニングを行った。Q 学習において

は空間をグリッドにしたことによっておよそ 20 回での収束という結果となった。mMLDA について、カテゴリ毎にトピックを抽出し、それらから統合概念を抽出した。概念の種類が 4 つと数が少ないため 95%の精度が出されていると考えられる。マルコフ性を考慮したバイグラムによる行動プランニングであるが、経験のみによる行動計画であるため、観測情報からの統合概念、および POMDP を実現する行動系列のプランニングに今後拡張する必要性がある。

4. おわりに

本研究ではロボットの知能化を目的とした、強化学習を用いた片付けに関する行動概念獲得を目指している。今回の実験は全てシミュレーションにとどまっているため、実機を用いた実験に今後取り組む。また、概念として行動を獲得することにより、行動条件に汎用性を持たせることが可能である。そのため、物体の色だけが異なる場合、形だけが異なる場合に関して概念の選択ができるのかを実験することも本研究の課題と考えられる。Q 学習による片付け行動獲得は、関節角単位などのより細かいレベルでの行動獲得が本来の課題に沿った形となっている。その点に関しても考慮していきたい。

謝辞

本研究は、JSPS 科研費 26280096 の助成を受けて実施した。

参考文献

- [1] Watkins, C.J.C.H., Learning from Delayed Rewards. PhD thesis, Cambridge University, Cambridge, England. 1989.
- [2] 浅田稔, 野田彰一, 俵積田健, 細田耕, 視覚に基づく強化学習によるロボットの行動獲得, 日本ロボット学会誌, Vol.13, No.1, pp.68-74, 1995.
- [3] 浅田稔, 野田彰一, 細田耕, ロボットの行動獲得のための状態空間の自律的構成, 日本ロボット学会誌, Vol.15, No.6, pp.886-892, 1997.
- [4] アッタミミ, ムハンマド, 阿部, 中村, 船越, 長井, 多層マルチモーダル LDA を用いた人の動きと物体の統合概念の形成, 日本ロボット学会誌, Vol.32, no.8, pp89-100, 2014.
- [5] 長井隆行, 中村友昭, アッタミミ ムハンマド, 持橋大地, 小林一郎, 麻生英樹, 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, 2F4-OS-01a-7, 2015.
- [6] T. Nakamura et al., "Grounding of Word Meanings in Multimodal Concepts Using LDA", in Proc. of IROS 2009, pp.3943-3948, 2009.
- [7] J. Pajarinen and V. Kyrki, "Robotic manipulation of multiple objects as a POMDP," Artificial Intelligence, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0004370215000570>.
- [8] J.D. Williams and S. Young. Partially observable markov decision processes for spoken dialog systems. Computer Speech & Language, Vol.21, No.2, pp393-422, 2007.