3P-07

強化学習を用いた物体の片付けに関する概念獲得への取り組み

恒川英里 † Muhammad Attamimi§ 小林一郎 † 長井隆行 § 中村友昭 §
麻生英樹 ‡ 持橋大地 ¶
† お茶の水女子大学 ‡ 産業技術総合研究所 § 電気通信大学 ¶ 統計数理研究所

1 はじめに

将来，ロボットが家庭に入ってきて人と共存した生活を送ることが予想される．そのためには，日常生活において人が行う行動の価値を理解し，自らの知識としてその行為の概念を獲得する必要が生じると考えられる．そこで，本研究では，適応的に学習し，行動を制御できるという利点のある強化学習を用いて，ロボットの行動知識獲得を目指す．本研究では物体を適切な場所に移動させるという行為を通じて,「片付けをする」という概念の獲得を実現することを目指す．

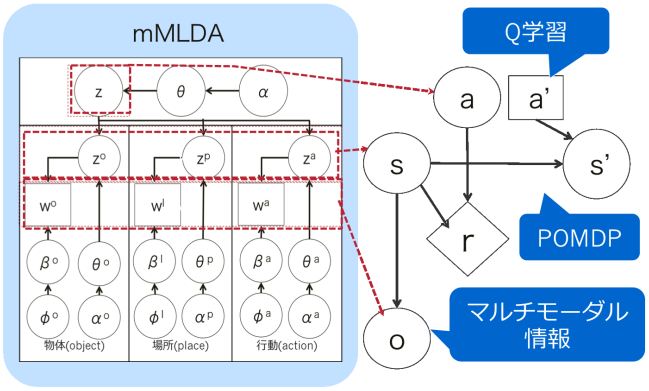

図 1: 研究概要図

2 マルチモーダル情報を用いた行動概念の獲得

2.1 概要

図 1 に研究概要図を示す．ロボットは観測するマルチモーダル情報に基づき，片付け対象となる物体に対してどの場所に置くべきかという行動概念を学習する．そして，学習した行動概念の系列を作り出すことにより，概念に基づくプランニングが可能になることを目指す．本研究では，マルチモーダル情報からの概念獲得に多層マルチモーダル LDA[1] を用い，行動概念の系列によるタスク処理のプランニングの枠組みに部分観測マルコフ決定過程 (POMDP) を用い，行動の方策の獲得に Q 学習を用いるとする．以下，それぞれについて説明する．

2.2 部分観測マルコフ決定過程

先行研究 [2] の提案手法に基づき，多層マルチモーダル LDA を通じて獲得された行動に対する状態および概念の情報を時系列に連結し，ロボットが最適行動系列をプランニングする際に POMDP を用いる．多層マルチモーダル LDA を用いた学習の際は下位概念で抽出されたそれぞれのトピックを状態，上位概念を行動として，プランニングを進めることとなる．図 1 の a は概念を示しており，a' は実際の行動を表している [2].

2.3 多層マルチモーダル LDA

多層マルチモーダル LDA(mMLDA) は下位層に下位概念を表現するマルチモーダル LDA(MLDA) を置き，上位層にそれらを統合する MLDA から成る．これにより，各々のカテゴリ分類を行うと同時に，それらの概念間の関係を教師なしで学習することができる．概念の獲得をこの枠組みを用いて行う．

2.4 Q 学習

本研究では，強化学習の枠組みにおいて最適な行動を学習する Q 学習によりロボットの行動知識を獲得する．観測したマルチモーダル情報から適当に状態空間を形成し，適当に行動していく過程で最適行動を導き出す方策を導出する枠組みとして用いる．また，Q 学習による行動価値の更新は，式 (1) によって示される．

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t)+\alpha[r_{t+1}+\gamma \max_a Q(s_{t+1},a)-Q(s_t,a_t)] \quad (1)$$

上式において，s は状況，a は行動，r_t は時刻 t における報酬，$Q(s,a)$ は累積報酬 $E\{R_t|s_t=s,a_t=a\}$ で表現される行動価値を表し，α は学習率，γ は割引率を表す．

An Approach to Concept Acquisition for the Act of Tidying Things up using Reinforcement Learning
Eri TSUNEKAWA†, Muhammad ATTAMIMI§, Ichiro KOBAYASHI†, Takayuki NAGAI§, Tomoaki NAKAMURA§, Hideki ASOH‡ Daichi MOCHIHASHI¶

3　予備実験

本研究では mMLDA と POMDP の枠組みでの状態空間と方策の並行的な学習であるが，今回は予備実験として，状態を定義した上での通常の MDP 用の Q 学習と mMLDA による状態空間の学習 (概念形成) とを別々に行う．

3.1　作業課題

テーブルの上に色が赤，青，形が丸，四角の合計 4 種類の積み木を設定した片付け場所に適切に置くという課題に対してマルチモーダル情報を観測し，状態空間を形成，Q 学習を行い，得られた方策結果から mMLDA による概念獲得を行った．

ロボットに取り付けられた頭部カメラ，ハンドカメラを用いて，テーブル上に置かれている色付きの物体を探索し，画像を取得する．画像中の物体に対して，画像処理ライブラリ OpenCV を用いた色認識および領域抽出による物体の認識を行う．HIRO はテーブル上の物体を発見した際に，その物体がどこにあるべきかを Q 学習によって学習し，最終的に物体に対して設定した正解の場所に置くという行動 (片付け) 知識を獲得する．

3.2　Q 学習による行動概念獲得

状態，行動，報酬を以下のように設定する．

- 状態：円形度，rgb 値，面積，場所，動き，の 5 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
- 報酬：置く，または，何もしないという行動をとった時の状態の物体が正解の場所にあれば，正の値の報酬，異なっていたら負の値の報酬，の 2 種類

続いて mMLDA を用いて概念獲得を行った．獲得の様子を図 2 に示す．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．

下位概念

- **物体：赤いかつ丸い，赤いかつ四角い，青いかつ丸い，青いかつ四角い，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

- **左上に赤く丸い物体を置く，左上に赤く四角い物体を置く，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 5 種類**

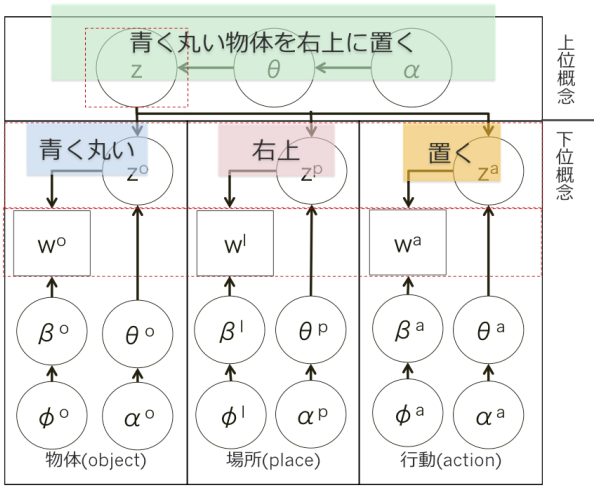

図 2: mMLDA を用いた概念獲得の様子

3.3　実験結果と考察

今回の実験において，得られたそれぞれのカテゴリにおける mMLDA の下位概念の予測精度は，それぞれ，物体，場所，行動は順に，63.6 %，100.0 %，89.8 %，であり，上位概念は 86.0 %であった．カテゴリ毎に抽出されたトピックに対して得られた上位概念は，マルチモーダル情報を用いた POMDP 学習を行うことのできる精度が得られたと考えられる．

次は，この結果を用いて POMDP 学習を行うことによるロボットの課題に対するプランニングに取り組む．

4　おわりに

本研究では，何も知識のない状態から環境に応じて観測を行い，状態空間を構成し，条件が異なる場合でも同じ結果を出せるような枠組みを目標としている．簡単な片付け課題に対して Q 学習で方策を獲得し，概念を形成する予備実験を行った．カテゴリ数を指定しない枠組みや，Q 学習で得た方策が行動系列として得られる場合，有効な行動のみを取り出すなどの発展性がある．

次の段階として，今回の結果を用いて POMDP 学習を行うことによってロボットが獲得した行動の概念を用いて課題達成のためのプランニングをさせることに取り組む．

参考文献

[1] **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.**

[2] **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，2015.**