The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

時系列マルチモーダル情報の分節・分類に基づく物体と動作の統合概念学習

Learning Integration Concept of Objects and Motions Based on Segmentation and Classification of Time Series Multimodal Information

布川 遼太郎 Ryotaro Nunokawa　宮澤 和貴 Kazuki Miyazawa　中村 友昭 Tomoaki Nakamura　長井 隆行 Takayuki Nagai　金子 正秀 Masahide Kaneko

電気通信大学
The University of Electro-Communications

Humans acquire various kinds of information with minimal cognitive load by using concepts and this capability makes it possible to deal with unknown events. Therefore, we consider that concepts also make it possible for robots to predict unobserved information from observable one. We define concepts as categories formed by clustering the perceptual information. In this paper, we propose a method for estimating a temporal segments of multimodal information based on co-occurrence relationship among modalities and classifying them into categories in an unsupervised manner. In the experiment, in which object images and robot's motions obtained by robot's manipulating objects were used, concepts that represents the co-occurrence relationship between objects and behaviors were formed.

1. 前書き

概念は人間の認知において重要な役割を担っている．人は概念を用いることで，知覚情報から最小限の認知的処理で様々な情報を獲得し [Rosch 99]，未知の事象にも対応している．Medin は概念の持つ 7 つの機能として，分類，理解と説明，予測，推理，コミュニケーション，概念の組み合わせ，事例の検索と生成を挙げている [Medin 85]．特に分類の機能は他の機能の基礎となるものとされており，認知心理学において研究されてきた．例えば認知科学の分野では，人は物体を見るとその物体の触感や動作などの関連情報を想像する“embodied simulation”という考え方があり [Barsalou 99]，この考え方を支援する多くの知見が存在している [Bergen 12]．これは人が“embodied simulation”を行うことで可観測情報から自身の経験に基づいて形成された概念を用いて未観測情報を予測していると言える．このことからロボットも概念を用いることによって，一部の可観測情報から未観測情報を予測することが可能となり，様々な事物に対して柔軟に対応することが可能となると考えられる．

そこで我々は，文献 [Miyazawa 16]，[Attamimi 14] においてロボットが物体に対して何らかの操作を行うことで，物体と動作の概念を統合した上位概念を獲得する手法を提案している．文献 [Miyazawa 16] で提案されたモデルでは各下位概念の関係性を表現する上位概念を学習することで，特定の物体に対する実行可能な操作を予測可能にしている．このモデルでは，下位概念の共起性に基づいて上位概念が形成される．しかし，実際にロボットが取得するセンサ情報は時系列情報であり，これらをクラスタリングした複数の下位概念の発生には時間的ズレが生じ，必ずしも共起するわけではない．そのため，ロボットは下位概念のどこからどこまでを共起としてとらえるかを推定する必要がある．本稿では概念を，知覚した情報をクラスタリングすることで形成されるカテゴリと定義し，ロボットが環境とのインタラクションを通じて取得可能な複数の時系列センサ情報を用いて概念形成を行う．そのために，隠れセミマルコフモデルを用いてセンサ情報を分節・分類することで，1 つの概念によって表現されるセンサ情報の時間的な範囲と，そのカテゴリを教師なしで推定する手法を提案する．

マルチモーダル情報を用いた概念獲得に関する研究として文献 [Mangin 15] がある．Mangin らは，概念を高次元のマルチモーダル情報を低次元で表現したパラメータであると定義している．そして，マルチモーダル情報は複数の知覚情報が相互に関係を持っているという考えに基づいて，複数の知覚情報を関連付けたマルチモーダル概念を獲得する手法を提案している．この手法では獲得したマルチモーダル概念を用いることで，可観測情報から未観測情報を予測することを可能にしている．しかし，この研究では文献 [Miyazawa 16] と同様に共起性に基づいて概念が形成されており，マルチモーダル情報の発生の時間的なズレは考慮されていない．

2. 提案手法

2.1 統合概念モデル

図 1 は提案手法のグラフィカルモデルであり，下位概念モデルにおいて観測情報 o_1 と o_2 の分節・分類によって下位概念を形成し，上位概念モデルで上位概念 z_i が下位概念 z_1^* と z_2^* の共起関係を表している．z_i では，下位概念の共起関係と，上位概念で表される下位概念の時間的な範囲を同時に推定する．例えば，観測情報 o_1，o_2 として物体画像とロボットアームの関節角度が観測されたとする．下位概念モデルにおいて o_1 と o_2 を分節化・分類することで物体概念・動作概念といった下位概念のカテゴリ系列を形成する．これらの下位概念のカテゴリ系列を上位概念モデルに送信し，1 つの上位概念で表現される下位概念の時間的な範囲を推定することで各物体と動作の対応関係を表現するような上位概念が形成される．すなわち，提案モデルを用いて学習を行うことでどの物体に対してどのような動作が実行可能であるかを表す上位概念を獲得することが可能となる．

2.2 時系列センサ情報の分節・分類

分節化に関する研究の一つに，ある文字列が与えられた時，これを意味の持つ単語に分割する形態素解析という手法がある．本稿では時系列センサ情報の分節・分類を，形態素解析における文字列をロボットが取得した時系列センサ情報に置き換えた問題と考える．この問題は，系列 $\boldsymbol{s}$ が与えられたとして，これを分割することで求められる単位系列 $\boldsymbol{x} = x_1, x_2$

連絡先: 布川遼太郎，電気通信大学知能機械工学科，東京都調布市調布ヶ丘 1-5-1，n1412098@edu.cc.uec.ac.jp

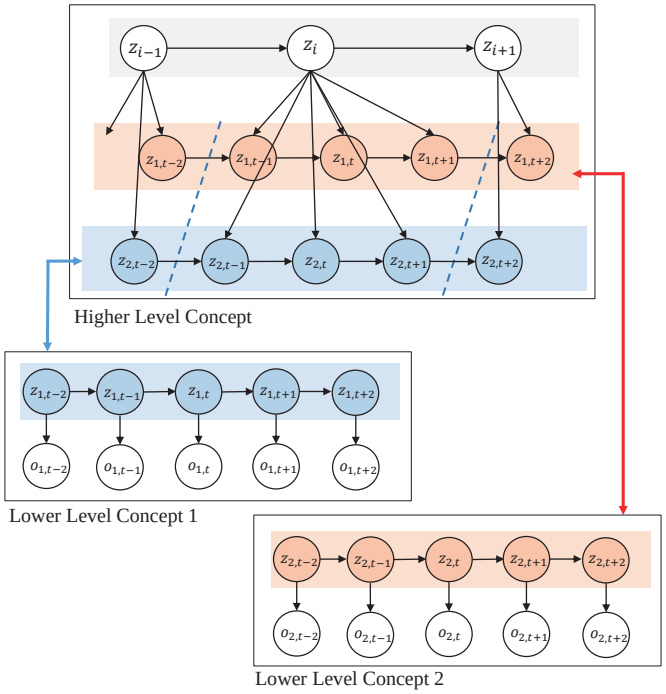

図 1: 時系列情報の分節・分類に基づく統合概念モデル

$, \cdots, x_N$ が生成される確率 $P(\boldsymbol{x}|\boldsymbol{s})$ を最大化する $\overline{\boldsymbol{x}}$ を求める問題に定式化される．

$$\overline{\boldsymbol{x}} = \mathrm{argmax} P(\boldsymbol{x}|\boldsymbol{s}) \tag{1}$$

ここで，$P(\boldsymbol{x}|\boldsymbol{s})$ を最大化することはできるだけ少ない単位系列数であり，かつ各単位系列の発生確率が高くなるような分節位置を推定することと考えられる．このような，時系列センサ情報の分節位置の推定は隠れセミマルコフモデル (HSMM) によって定式化することができる．パラメータ推定には 1 つの観測系列内の単位系列とそのクラスをまとめてサンプリングする Blocked Gibbs Sampler を用いる．また，Blocked Gibbs Sampler の計算を行う際には計算の効率化のために Forward filtering-backward sampling を用いる．

2.3 上位概念のカテゴリ系列の推定

上位概念のカテゴリ系列 $\boldsymbol{z}$ は次式で推定することができる．

$$\boldsymbol{z} \sim \iint P(\boldsymbol{z}|\boldsymbol{z_1}, \boldsymbol{z_2}) P(\boldsymbol{z_1}|\boldsymbol{o_1}) P(\boldsymbol{z_2}|\boldsymbol{o_2}) d\boldsymbol{z_1} d\boldsymbol{z_2} \tag{2}$$

ここで，$\boldsymbol{z}, \boldsymbol{z_1}, \boldsymbol{z_2}, \boldsymbol{o_1}, \boldsymbol{o_2}$ はそれぞれ $z_i, z_{1,t}, z_{2,t}, o_{1,t}, o_{2,t}$ のベクトルである．しかし，$\boldsymbol{z_1}, \boldsymbol{z_2}$ の積分は下位概念モデルから独立した上位概念モデルのみでは直接計算することが困難であるため，近似することで計算する．まず下位概念モデルで，下位カテゴリ系列 $\boldsymbol{z_1^*}, \boldsymbol{z_2^*}$ を推定する．

$$\boldsymbol{z_1^*} \sim P(\boldsymbol{z_1}|\boldsymbol{o_1}) \tag{3}$$

$$\boldsymbol{z_2^*} \sim P(\boldsymbol{z_2}|\boldsymbol{o_2}) \tag{4}$$

これにより上位概念モデルでは，推定された $\boldsymbol{z_1^*}$ と $\boldsymbol{z_2^*}$ を用いて式 (5) を Forward filtering-backward sampling によって計算することができる．

$$\boldsymbol{z}^* \sim P(\boldsymbol{z}|\boldsymbol{z_1^*}, \boldsymbol{z_2^*}) \tag{5}$$

Algorithm 1 上位カテゴリ系列の推定

```
下位カテゴリ系列の推定

// Forward filtering
for t = 1 to T do
    for k = 1 to K do
        for z = 1 to Z do
            α[t][k][z] を計算
        end for
    end for
end for

// Backward sampling
t = T, j = 1
while t > 0 do
    k, z ~ p(x_j | s_{t-k:t}) α[t][k][z]
    x_j = s_{t-k:t}
    z_j = z
    t = t - k
    j = j + 1
end while
return (x_{J_n}, x_{J_n - 1}, ..., x_1), (z_{J_n}, z_{J_n - 1}, ..., z_1)
```

Forward filtering では，動的計画法を用いて，時刻 t を終点とする分節長 k の単位系列 $\boldsymbol{z}_{*,t-k:t}$ が，クラス z になる確率 $\alpha[t][k][z]$ をあらゆる可能性について計算する．

$$\alpha[t][k][z] = P(\boldsymbol{z}_{1,t-k:t}|z) P(\boldsymbol{z}_{2,t-k:t}|z) \sum_{k',z'} \alpha[t-k][k'][z'] \tag{6}$$

式 (6) における $P(\boldsymbol{z}_{1,t-k:t}|z)$ と $P(\boldsymbol{z}_{2,t-k:t}|z)$ はクラス z から分節長 k の単位系列が出力される確率を表しており，次式のように表される．

$$P(\boldsymbol{z}_{m,t-k:t}|z) = \prod_{t'=t-k}^{t} \frac{N_{z,z_{m,t'}} + \alpha}{N_z + Z\alpha} \tag{7}$$

ただし，$N_{z,z_{m,t'}}$ は下位概念モデル m において下位カテゴリが $z_{m,t'}$ となったもののうち上位概念 z に分類された個数を，N_z は上位カテゴリ z に分類された下位カテゴリの総数を表している．α 及び Z はそれぞれディリクレ事前分布のパラメータ，カテゴリ数を表す．Backward sampling では，計算された $\alpha[t][k][z]$ からサンプリングすることで系列の分節位置を決定する．以上のアルゴリズムを，Algorithm 1 に示す．K は最大の分節長である．

3. ロボットによる概念学習実験

3.1 実験設定

前章で説明したモデルを用いて，実際にロボットが物体を操作し得られる時系列マルチモーダル情報を分節・分類することで概念の学習を行った．下位概念としては物体概念と動作概念を用いた．図 2 のようにロボットはマルチモーダル情報を取得する．操作毎にロボットの頭部に取り付けた Kinect によって操作対象の物体の画像を取得し，LDA[Blei 03] を用いて分類することで得られる物体 ID を物体概念とした．また，物体操作時のロボットアームおよびハンドの関節角度を GP-HSMM[Nakamura 17] によって分節することで得られる動作プリミティブを動作概念とした．すなわち，ロボットはどの物体に対してどのような動作を実行可能か表す上位概念を学習することになる．実験には図 3 に示した物体を用い，表 1 に示す動作を各物体に対して行った．実際のロボットによるマルチモーダル情報の取得は以下の流れで行った．

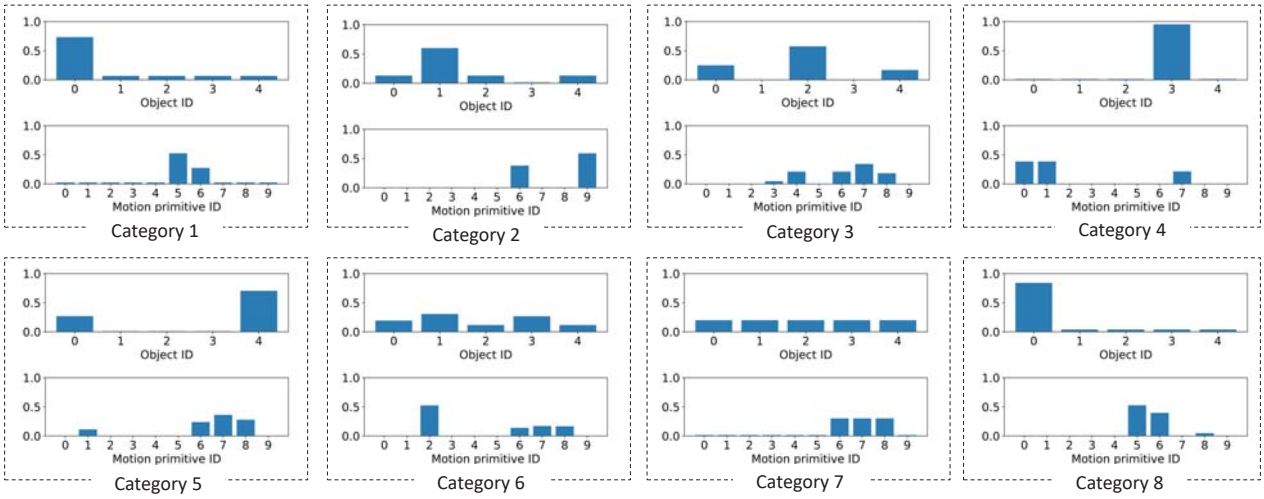

図 5: 各カテゴリにおける物体と動作プリミティブ系列の発生確率

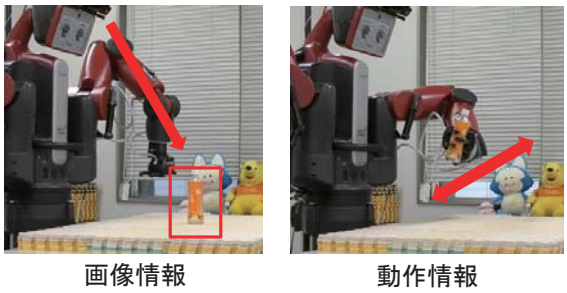

図 2: ロボットによるマルチモーダル情報の取得

表 1: 物体と動作の対応

物体 (ID)	動作内容	動作プリミティブ系列
ぬいぐるみ (0)	握る・なでる	7,8,7,8,6・5,5,6
ボタン (1)	押す	9,9,6
おもちゃ(2)	左に置く	7,3,6
ペットボトル (3)	振る	7,1,0,1,0,7,8,6
スポンジ (4)	持ち上げる	7,1,7,8,6
おもちゃ・スポンジ	擦る	7,4,7,4,8,6
全物体	観察する	7,2,2,2,8,6

1. ロボットアームを人が操作し，各物体に対して行う操作の関節角情報を取得する．
2. GPHSMM を用いて関節角情報から動作プリミティブを生成する．
3. 以下の操作を各物体に対して繰り返し行う
   - **(a)** ロボットが机上の物体の画像情報を取得する．
   - **(b)** 物体に対応する動作を動作プリミティブを組み合わせて生成し，ロボットに行わせる．
   - **(c)** 取得した画像情報と動作情報を保存する．
   - **(d)** 人が新たに物体を机に置き (a) に戻る

図 4 が取得した画像と関節角の 2 つの連続的な情報を分節・分類することで得られる物体 ID と動作プリミティブの時系列データの一部である．図 4 において横軸はタイムステップを，縦軸は物体 ID 及び動作プリミティブの ID を表しており，縦軸の値が -1 である状態は物体については物体情報を観測していない状態を，動作に関してはロボットが動作を行っていない状態であることを表している．破線で囲まれた範囲が一つの上位概念によって表現される範囲を表しており，これを教師なしで推定する．

図 3: 実験に用いた物体

3.2 結果

図 5 が実際に分節・分類することで得られた，各カテゴリにおける物体 ID と動作プリミティブ系列の発生確率である．Category 1 において，物体 0 に対して正解である動作プリミティブ 5・6 によって構成される動作が行われることが学習されている．同様に Category 2・3・5・6 においても，各物体に対応する実行可能な動作が正しく学習されていることがわかる．

一方で Category 1 と Category 8 の両方で，物体 0 に対してプリミティブ 5・6 で構成される動作が実行可能であることが学習されており，正解であるプリミティブ 6・7・8 で構成さ

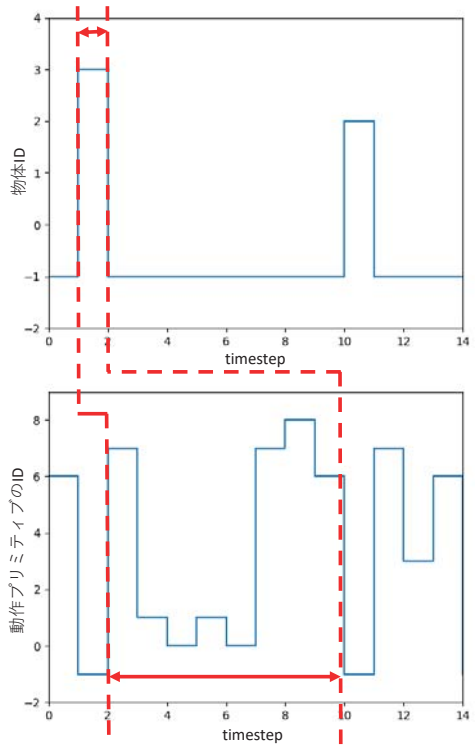

図 4: 取得された物体 ID(上) と動作プリミティブ (下)

れる動作が実行可能であることを表現する結果は得られなかった．これは，物体 0 に対して実行可能なプリミティブ 6・7・8 で構成される動作（7,8,7,8,6）と，物体 4 に対して実行可能なプリミティブ 1・6・7・8 によって構成される動作 (7,1,7,8,6) が非常に似ているため，学習が正しくできなかったことが原因であると推察される．その結果，Category 5 に，プリミティブ 1・6・7・8 によって構成される動作が実行可能な物体として，正解である物体 4 に加えて正解でない物体 0 が学習結果として表れている．

Category 4 では物体 3 に対して，想定した動作プリミティブ 0・1・6・7・8 で構成される動作が実行可能であることをある程度学習できているものの，動作プリミティブ 6・8 の発生頻度が非常に低くなってしまっていた．これは想定した動作において動作プリミティブ 6・8 の発生回数が 0・1・7 に比べ少なく，また，8 種類の操作のうち 5 種類の操作にプリミティブ 6・7・8 が含まれているため，プリミティブ 0・1・6・7・8 で構成される動作がプリミティブ 6・7・8 で構成される動作と同じカテゴリに一部含まれてしまったことが原因であると考えられる．

物体 2 に対して実行可能なプリミティブ 3・6・7 で構成される動作が実行可能であることが学習されていない．そこで，Category 3 を見ると動作プリミティブ 3 が発生していることがわかる．このことから，プリミティブ 3・6・7 で構成される動作がプリミティブ 4・6・7・8 で構成される動作と同じカテゴリに含まれてしまい，学習が正しく行えなかったと考えられる．

実験結果より 8 つのカテゴリ中，5 つのカテゴリについて学習を正しく行うことができることが確認された．また，正しく学習が行えなかった 3 つのカテゴリのうち 2 つのカテゴリではある程度学習が正しく行えており，提案手法を用いることでロボットが時系列マルチモーダル情報から概念を学習可能であるとわかった．

4. 結び

本稿では，ロボットが物体を操作し得られる時系列マルチモーダル情報を分節・分類することによって概念を学習する手法を提案した．実際にロボットを用いた実験を行い，提案手法によって上位概念の学習が可能であることを示した．今後の課題として，今回の実験結果で見られたような，似ている動作が行われた場合に学習が正しく行えない点を改善することや，上位概念モデルと下位概念モデルが相互に影響しあい，各モデルのパラメータ推定を行うモデルの実装などが挙げられる．

参考文献

[Rosch 99] Rosch, B.: Principles of Categorization, Concepts: core readings, pp.189-206 (1999)

[Medin 85] Eysenck, M. W., Ellis, A., Hunt, E. and Johnson-Laird, P. N.: The Blackwell dictionary of cognitive psychology, Blackwell Pub (1985)

[Barsalou 99] Barsalou, L. W.:Perceptual symbol system, Behavioral and Brain Sciences, vol. 22, pp.277-660 (1999)

[Bergen 12] Bergen, B. K.: Louder Than Words, The New Science of How the Mind Makes Meaning. Basic Books (2012)

[Mangin 15] Mangin, O., Filliat, D., Bosch, L. T., Oudeyer, P. Y.: MCA-NMF: Multimodal Concept Acquisition with Non-Negative Matrix Factorization, PLoS ONE 10(10): e0140732. doi:10.1371/journal. pone.0140732 (2015)

[Miyazawa 16] 宮澤 和貴, 青木 達哉, 日永田 智絵, Mata, D., 岩田 健輔, 中村 友昭, 長井 隆行：多層マルチモーダル LDA を用いた概念形成と強化学習，計測自動制御学会システム・情報部門 学術講演会, O1-2 (2016)

[Attamimi 14] Attamimi, M., Fadlil, M., Abe, K., Nakamura, T., Funakoshi, K., and Nagai, T.: Integration of various concepts and grounding of word meanings multi-layered multimodal LDA for sentence generation, in IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.2194-2201 (2014)

[Blei 03] Blei, D. M., Ng, A. Y., and Jordan, M. I.: Latent Dirichlet Allocation, Journal of Machine Learning Research, Vol. 3, pp.993-1022 (2003)

[Nakamura 17] Nakamura, T., Takayuki, N., Daichi, M., Ichiro K., Hideki, A., Masahide, K.: Segmenting Continuous Motions with Hidden Semi-markov Models and Gaussian Processes, Front. Neurorobot, doi: 10.3389/fnbot.2017.00067 (2017)