

複数概念の時間的分節化に基づく上位概念の学習

Formation of Higher-Level Concept Based on Temporal Segmentation of Multiple Concepts

中村 友昭*¹ 宮澤 和貴*¹ 青木 達哉*¹ 長井 隆行*¹ 金子 正秀*¹
Tomoaki Nakamura Kazuki Miyazawa Tatsuya Aoki Takayuki Nagai Masahide Kaneko

*¹電気通信大学情報理工学研究所
The University of Electro-Communications

There are various concepts in the real environment, and higher-level concepts can be formed by combining those concepts. We have proposed a probabilistic model that enables robots to learn such higher-level concepts. The basic idea to form the higher-level concepts is co-occurrence of lower-level concepts. For instance, when we “drink” something using “cup” in the “kitchen”, the motion, the object and the place can be recognized by using the corresponding concepts. Furthermore, higher-level concepts can be formed by combining these concepts, and enables us to predict among lower-level concepts mutually. However, the lower-level concepts are not always co-occurred at the same time, and there are time lags between occurrence of the lower-level concepts. In this paper, we propose a probabilistic model that segments and classifies the lower-level concepts considering such time lags, and form the higher-level concepts.

1. はじめに

概念は人間の認知のあらゆる局面で重要な役割を果たしている。人間は概念を形成することで、経験した物事を全て参照することなく、必要最小限の認知的処理によって、より多くの情報を得ることができる。さらに概念を用いることで、直接的には観測できない未観測の情報を予測することもでき、この予測により人は未知の物事に対しても柔軟に対応している。我々は、概念とは知覚した情報をクラスタリングしたカテゴリであり、概念を通して予測こそが理解であると定義し、ロボットが他者や環境とのインタラクションを通じて概念を獲得する手法を提案してきた。文献 [Nishihara 16] では、ロボットが物体を観察して、触れて、振ること得られるマルチモーダル情報と、他者とのインタラクションによって得られる言語情報から、ロボットが物体概念と言語的知識を獲得可能であることを示した。また、文献 [中村 16] では、ロボットが観測した人の動作を分節・分類することで、動作概念が獲得できることを示した。さらに、これら以外にも、谷口らは場所の概念獲得手法 [Taniguchi 16] を、Zhixiang らは相対的な概念獲得手法 [Gu 16] を提案している。このように、様々な概念モデルが提案されているが、ロボットが何かしらのタスクを達成するために、これらの概念を相互に利用する必要がある。そこで我々は、様々な概念を統合した上位概念 [Attamimi 14, 宮澤 16] が獲得可能な手法を提案した。この手法では、物体と動作の概念の関係性を表現した上位の概念を学習することで、特定の物体に対して実行可能な動作を予測することができる。このような概念間関係を学習することで、ロボットはより複雑なタスクを実行することが可能となる。

しかし、このモデルには2つの問題が存在している。1つはモデル構造が事前に決まっており、新たな概念モデルを容易に導入することができない点である。そこで、本稿ではこの問題に対して、下位の複数の概念モデルと、上位の概念モデルを独立したモジュールに分解し、それらが通信することで概念を形成する枠組みを提案する。それぞれの概念モデルを独立したモジュールと考えることで、モデル構造をあらかじめ固定するこ

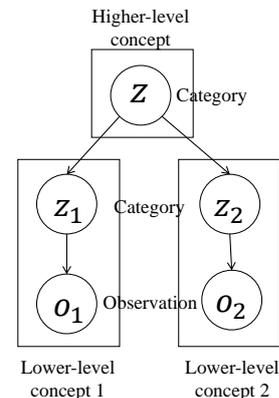


図 1: 統合概念モデル

となく、モジュールの追加や入れ替えを容易に行うことができる。また、2つ目の問題は、これまでの上位概念は、複数の下位概念の共起性に基づいて形成されている点である。すなわち下位の概念へは、1つの上位概念で表現される情報が、同時に入力されることが仮定されている。しかし、実際にはロボットが物体を見ることで物体概念へと情報が入力され、その後ロボットが動作することで動作情報が動作概念へと入力される。すなわち、実際には時間差を伴って入力されるマルチモーダル情報のどこからどこまでを1つの上位概念で表現するべきかを判断する必要がある。そこで本稿では、マルチモーダル情報を分節・分類することで上位概念を形成可能な手法を提案する。

2. 統合概念モデル

図 1 が下位概念を上位概念によって統合した概念モデルのグラフィカルモデルである。このモデルでは、 z が上位概念を表すカテゴリであり、この上位概念から下位の概念を表現する

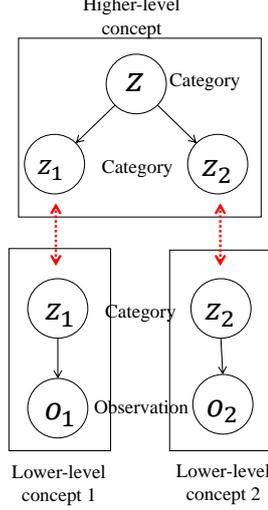


図 2: 上位概念と下位概念のモジュール化

カテゴリ z_1 と z_2 が生成される。

$$z \sim P(z) \quad (1)$$

$$z_1 \sim P(z_1|z) \quad (2)$$

$$z_2 \sim P(z_2|z) \quad (3)$$

すなわち、 z が下位のカテゴリ z_1 と z_2 の共起関係を表現している。さらに、下位カテゴリからは観測 o_1 と o_2 が生成される生成モデルである。

$$o_1 \sim P(o_1|z_1) \quad (4)$$

$$o_2 \sim P(o_2|z_2) \quad (5)$$

このモデルの下位概念モデルの追加や入れ替えが容易になるよう、それぞれを独立したモジュールとして考え、図 2 のように分解する。このモデルにおいて、個々のモデルを独立して学習することも可能であるが、我々のこれまでの実験により独立した学習に比べ、相互に影響を与えることで個々のモデルの性能が向上することがわかっている [Attamimi 14]。そこで、本稿でも上位概念と下位概念が通信しあうことで相互に影響を与えながら学習することを考える。ただし、学習は Collapsed Gibbs Sampler を用いることを想定している。

まず、上位カテゴリ z を推定するためには、観測 o_1 と o_2 から z が発生する確率を計算する必要がある。

$$z \sim P(z|o_1, o_2) \quad (6)$$

$$\propto \int \int P(z|z_1)P(z|z_2)P(z_1|o_1)P(z_2|o_2)dz_1dz_2 \quad (7)$$

この分布の計算に必要な $P(z|z_1)$ と $P(z|z_2)$ は、上位の概念モジュールがすでに持っている情報であるため、必要となるのは $P(z_1|o_1)$ と $P(z_2|o_2)$ であり、そのパラメータを下位概念から上位概念に送信することで上位概念の学習が可能となる。 $P(z_1|o_1)$ と $P(z_2|o_2)$ は多項分布となるので、そのパラメータは下位カテゴリの要素数分の実数ベクトルであるため容易に送信することができる。

次に、下位カテゴリ z_1 を学習することを考える。 z_1 を推定

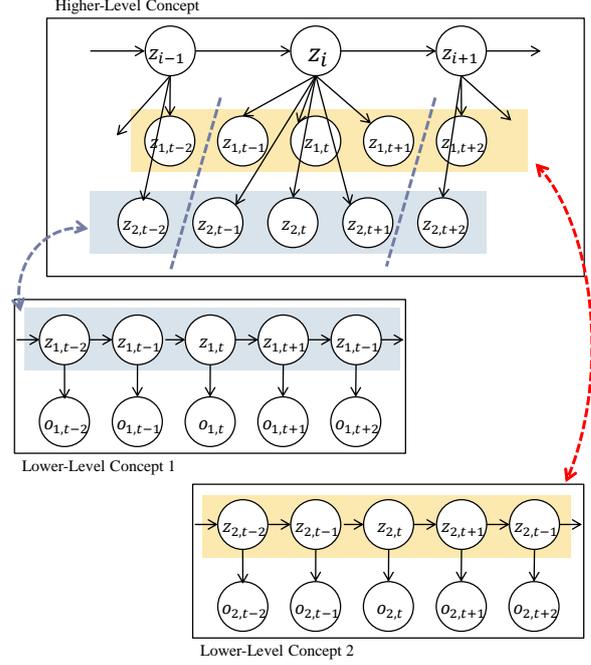


図 3: 下位概念の分節・分類に基づく統合概念モデル

するためには、以下の確率分布を計算する必要がある。

$$z_1 \sim P(z_1|o_1, o_2) \quad (8)$$

$$\propto \int P(z_1|o_1)P(z_1|z)P(z|z_2)P(z_2|o_2)dz_2 \quad (9)$$

$$= P(z_1|o_1)P(z_1|o_2) \quad (10)$$

この確率分布を計算するにあたり、 $P(z_1|z)$ と $P(z|z_2)$ は上位概念モジュールのパラメータであり、 $P(z_2|o_2)$ は前述の上位概念の学習の際に下位概念モジュールから送られている情報である。すなわち、上位概念モジュールによって $P(z_1|o_2)$ が計算可能であり、これを下位概念モジュールに送ることで、 $P(z_1|o_1, o_2)$ が計算可能となる。 $P(z_1|o_2)$ は多項分布であるため、そのパラメータはカテゴリの要素数分の実数ベクトルであり容易に送信することができる。

以上のように、上位概念モジュールと下位概念モジュールが通信することで相互に影響しあいながら、それぞれのモデルのパラメータを推定することが可能となる。

3. 下位概念の分節・分類に基づく上位概念の形成

前節では、下位の概念の共起性を学習することで上位の概念を形成した。しかし、下位概念は必ずしも共起するわけではなく、時間的なずれが生じる。例えば、物体に対する動作を考えると、ロボットは物体を見ることで物体概念を使用し物体を認識することができ、その後動作概念を利用することで、その物体に対して何かしらの行動をとることができる。この時、物体と動作が同時に共起しているわけではなく、時間にずれが生じることになる。すなわち、ロボットはどこからどこまでを共起として捉えるべきかを推定する必要がある。これは分節化の問題であり、ロボットは入力されるマルチモーダル情報を分節・分類し下位概念と上位概念を形成する必要がある。

そこで、図 2 を時間展開したモデルである図 3 のモデルを考える。これは、図 2 と同様に下位の概念の $z_{1,t}$ と $z_{2,t}$ の共起関係を上位概念 z_i で表現したモデルである。この時、 z_i では単に共起した下位概念を学習するだけでなく、どこからどこまでの範囲が z_i で表現可能なのかも同時に推定する。本稿では、隠れセミマルコフモデル (HSMM) を用いることで、この上位概念を学習する。

まず、上位概念のカテゴリ系列 z は次式により推定することができる。

$$z \sim \int \int P(z|z_1, z_2)P(z_1|o_1)P(z_2|o_2)dz_1dz_2 \quad (11)$$

ただし、 z, z_1, z_2, o_1, 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、式 (11) における、 z_1 と z_2 による積分は下位概念のあらゆる組み合わせで計算する必要があり、下位概念モデルから独立した上位概念モデルだけでは計算することができない。そこで、まず下位概念モデルによって下位カテゴリの系列 z_1^*, z_2^* を推定する。

$$z_1^* \sim P(z_1|o_1) \quad (12)$$

$$z_2^* \sim P(z_2|o_2) \quad (13)$$

推定された下位カテゴリを固定し、上位概念のカテゴリ z^* を推定する。

$$z^* \sim P(z|z_1^*, z_2^*) \quad (14)$$

これにより下位概念モジュールは推定された z_1^* と z_2^* を上位概念モジュールへ送信することで、式 (14) を Forward filtering-backward sampling によって計算することができる。Forward filtering では、動的計画方を用いて、時刻 t を終点とした長さ k の下位概念の系列 $z_{*,t-k:t}$ が上位概念 z になる確率を以下のように計算する。

$$\alpha[t][k][z] = P(z_{1,t-k:t}|z)P(z_{2,t-k:t}|z) \sum_{k', z'} \alpha[t-k][k'][z'] \quad (15)$$

この確率に従い、後ろ側からサンプリングすることで、 z^* を決定することができる。

次に、下位カテゴリの系列 z_1 を推定する場合を考える。

$$z_1 \sim \int \int P(z_1|z)P(z_1|o_1)P(z|z_2)P(z_2|o_2)dzdz_2 \quad (16)$$

ここでも、 z と z_2 での積分があり、独立したモデルでは容易に計算することができない。そこで、ここでも推定された上位概念のカテゴリ z^* を用いることで、次式により下位カテゴリの系列を決定することができる。

$$z_1 \sim P(z_1|z^*)P(z_1|o_1) \quad (17)$$

すなわち、上位概念モジュールは推定された z^* から、 z_1 が出力される確率分布を下位概念モジュールに送信することになる。この確率分布は多項分布となるため、これまでと同様にそのパラメータを容易に送信することができる。

4. 実験

提案手法の有効性を確認するため、基礎的な実験を行った。



図 4: 実験に用いた物体

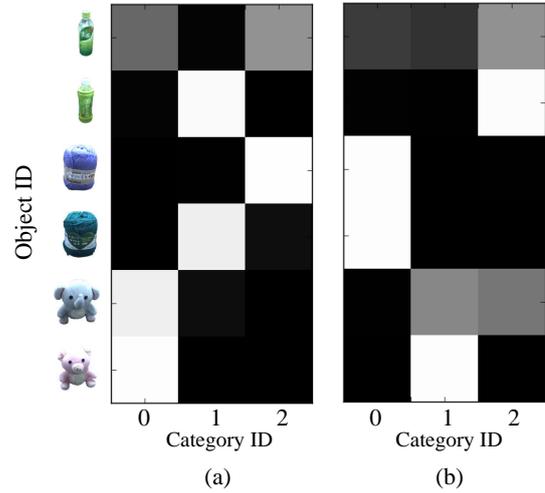


図 5: 画像の分類結果 (a) 上位概念からのフィードバックなし (b) 上位概念からのフィードバックあり

4.1 統合概念モデルによる物体概念の形成

まず、2. 節で述べたモデルを利用し、物体画像と言語から物体概念の形成実験を行った。上位概念の実装には、Multimodal Latent Dirichlet Allocation (MLDA) [Nakamura 12] を用い、下位概念の実装には LDA [Blei 03] を用いた。また、下位概念モデルと上位概念モデルの通信には Robot Operating System (ROS) のサービスを利用した。実験には図 4 に示した物体を利用し、これらの画像から局所特徴である KAZE 特徴量 [Alcantarilla 12] の Bag of Features を下位概念の学習に用いた。もう一つの下位概念の学習には、それぞれの物体の特徴を教示する教示文 (例: 「これ / わ / ぬいぐるみ / だよ」) の Bag of Words を用いた。すなわち、画像から形成される概念と単語から形成される概念が、上位概念を通して相互に影響し合うことで、学習されることになる。実験では、この相互の影響を確認するため、上位概念から下位概念へのフィードバックがあるものと、ないもので比較を行った。

図 5 が画像の分類結果であり、画像情報 o_1 が各カテゴリ z_1 に分類される確率を色の濃淡で表している。図 5(a) が上位概念からのフィードバックがない場合であり、一部の物体が正しく分類できていないことが分かる。一方、図 5(b) は上位概念からフィードバックがあり、上位概念を通して下位概念同士が影響を与えるモデルである。図 5(b) では、誤分類が修正され正しく分類できていることが分かる。以上のように、本稿で提案するフレームワークにより、下位の複数の概念同士が相互に影響を与えながら学習することができていることが分かる。

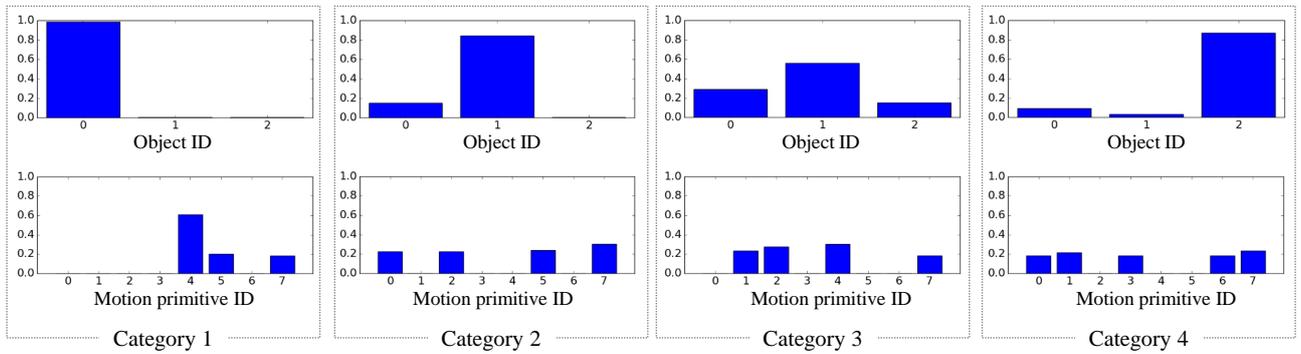


図 6: 上位概念に分節・分類された各カテゴリにおける物体と動作の発生確率

表 1: 物体と動作の対応

物体 (ID)	動作プリミティブ系列
ぬいぐるみ (0)	7,4,4,4,5
ペットボトル (1)	7,0,2,5
毛糸 (2)	7,0,1,3,6
全物体	7,4,2,1

4.2 下位概念の分節化に基づく上位概念形成

次に、3. 節で説明したモデルを利用し、下位概念を分節・分類することによる上位概念の形成を行った。実験では、文献 [宮澤 16] で用いたタスクの一部をシミュレーションしたものを使用した。このタスクでは、ロボットが物体を観察し下位概念を用いて物体を認識し、その物体と対応した動作を実行する。これを繰り返すことで連続的なマルチモーダル情報が取得される。動作は文献 [宮澤 16] と同様に、GP-HSMM [中村 16] を下位概念とすることで、動作のプリミティブの系列として表現される。例えば、ぬいぐるみに対する「握る」という動作は、 $7 \rightarrow 4 \rightarrow 4 \rightarrow 4 \rightarrow 5$ という動作プリミティブの系列として表現される。文献 [宮澤 16] では、ロボットが物体を観察し、その物体に対する適切な行動を強化学習によって学習すると同時に、物体と動作から形成される上位概念を学習している。しかし、ロボットが物体を観察してから、動作が終了するまでの範囲は人手で与えている。すなわち、どこからどこまでが 1 つの上位カテゴリで表現できるかが与えられており、分節化は考えられていない。そこで、本実験では図 4 の 3 種類の物体に対して、各物体に対する動作プリミティブの系列を表 1 のように擬似的に設定し、物体と動作の連続的な下位概念の系列を分節・分類し、上位概念の形成が可能か検証した。ただし、各物体に対して 1 つの動作系列の他に、本実験では全ての物体に対して実行可能な動作として $7 \rightarrow 4 \rightarrow 2 \rightarrow 1$ という動作系列を使用した。

図 6 が実際に上位概念で分節・分類することで得られた各カテゴリにおける、物体と動作プリミティブの発生確率である。Category 1 では、物体 0 に対して、 $7 \cdot 4 \cdot 5$ のプリミティブで構成される動作が実行可能であることが正しく学習されている。さらに Category 3 では、全ての物体に対して実行可能な動作が正しく表現されている。

5. まとめ

本稿では、下位概念の共起関係を捉えて上位概念を形成可能な統合概念モデルを、分解しモジュール化することで、概念モデルの追加・入れ替えが容易に行える手法を提案した。提案手法では、モジュール同士が通信することで、相互の概念モデルの精度を高めることが可能であり、実験によりその有効性を確認した。さらに、時間的に展開し、下位の概念系列を分節・

分類することで上位概念が学習可能な手法を提案した。基礎的な実験ではあるが、提案手法によって上位概念が学習可能であることが示された。

今後の課題として、実際のロボットから得られるマルチモーダル情報を利用した、より規模の大きい実験により提案手法の有効性を検証することがあげられる。さらに、今回提案した概念モデルの通信をより汎用的なものに拡張し、複数の確率モデルの統合が可能なフレームワークを構築することを考えている。

謝辞

本研究は JST CREST, JSPS 科研費 JP16H02835 の助成を受け実施したものである。

参考文献

- [Alcantarilla 12] Alcantarilla, P. F., Bartoli, A., and Davison, A. J.: KAZE Features, in *European Conference on Computer Vision*, pp. 214–227 (2012)
- [Attamimi 14] Attamimi, M., Fadlil, M., Abe, K., Nakamura, T., Funakoshi, K., and Nagai, T.: Integration of various concepts and grounding of word meanings using 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)
- [Gu 16] Zhixiang, G., 田口 亮, 服部 公亮, 保黒 政大, 梅崎 太造: フォン・ミーゼス分布モデルを用いた相対的な位置概念の学習, 計測自動制御学会 システム・情報部門 学術講演会, GS13-3 (2016)
- [Nakamura 12] Nakamura, T., Araki, T., Nagai, T., and Iwahashi, N.: Grounding of Word Meanings in LDA-Based Multimodal Concepts, *Advanced Robotics*, Vol. 25, pp. 2189–2206 (2012)
- [Nishihara 16] Nishihara, J., Nakamura, T., and Nagai, T.: Online Algorithm for Robots to Learn Object Concepts and Language Model, *IEEE Transactions on Cognitive and Developmental Systems*, Issue 99 (2016)
- [Taniguchi 16] Taniguchi, A., Taniguchi, T., and Inamura, T.: Spatial concept acquisition for a mobile robot that integrates self-localization and unsupervised word discovery from spoken sentences, *IEEE Transactions on Cognitive and Developmental Systems*, Vol. 8, No. 4, pp. 285–297 (2016)
- [宮澤 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)
- [中村 16] 中村 友昭, Muhammad, A., 長井 隆行, 持橋 大地, 小林 一郎, 麻生 英樹, 金子正秀: ガウス過程の隠れセミマルコフモデルに基づく身体動作の分節化, 人工知能学会全国大会, 103-5 (2016)