# 統合認知モデルによるロボットの概念・行動・言語の同時学習

Integrated Cognitive Model for Robot Learning of Concepts, Actions, and Language

宮澤 和貴 *1	青木 達哉 *1	堀井 隆斗 *1	長井 隆行 *1*2
Kazuki Miyazawa	Tatsuya Aoki	Takato Horii	Takayuki Nagai

\*<sup>1</sup>電気通信大学 \*<sup>2</sup>大阪大学 The University of Electro-Communications Osaka University

It is expected to realize an intelligent robot that coexists with us and supports our lives. In order to realize such a robot, it is necessary to act by understanding the surrounding environment and language in the real world. For such a complex understanding, multiple cognitive functions are required. The purpose of this study is to propose an integrated cognitive model for robots that simultaneously learn concepts, actions, and language. This can be achieved by integrating various modules. The proposed model consists of three modules: concepts formation using multimodal information, grammar learning, and reinforcement learning. We verify the potential of the proposed model through experiments using a real robot. We show that the proposed model enables robots to form concepts, make decision, and understand language.

# 1. はじめに

人間の生活を支援するような知的なロボットの登場が期待 されている.このようなロボットは、周囲の環境や言語を理解 して行動する必要がある.筆者らは、記号創発ロボティクスの 視点から言語を含むロボットによる実世界の理解について検討 を続けてきた [Taniguchi16].ここで理解とは予測することで あると考え、この予測を効率的に行うために概念形成を行う. 概念とは様々なセンサにより取得したマルチモーダル情報をカ テゴライズした過去の経験であると考える.中村らは、文章分 類で用いられる Latent Dirichlet Allocation(LDA)をマルチ モーダル情報に拡張したマルチモーダル LDA(MLDA)を提案 し、ロボットを用いた実験を行うことで、ロボットによる概念 形成の可能性を示した [Nakamura09].また、MLDAを拡張 し他のモジュールと結合することで、簡単な言語や行動を扱う ことができる [Attamimi16, Miyazawa17].

一方で近年の深層学習技術の発展により、ロボットの様々なタ スクにおける深層強化学習の有用性が示されている [Levine16, Gu16]. さらには、本研究と同様にマルチモーダルデータを用 いた学習や、言語と行動の同時学習も盛んに研究されている [Arandjelovi17, Hill17]. これらのアプローチでは、知能の一 部に焦点を当てモデル化することによりある機能単位の知能を 獲得している.しかし実世界のような複雑な環境で動作するロ ボットを作るには、概念形成や言語理解、行動決定などの個々 の認知機能が相互作用するような複雑な知能を考る必要があ る.また、深層学習による知能獲得はシミュレーションを用い たものが多く実際のロボットを用いた研究は少ない.

そこで本稿では、複数の認知モジュールを統合した統合認知 モデル(計算モデル)を構築し実際のロボットを用いて実装・ 評価することを目標とする.提案する統合認知モデルでは、概 念・行動・言語をロボット自身の経験から同時に学習する.こ のように、自身の経験からボトムアップに複数の認知機能を獲 得するモデルを構成することは、構成論的アプローチの観点か らも、人間の学習メカニズムを解明する一助となるため有意義 である.本研究では、提案モデルの有効性を検証するために、

連絡先: 宮澤 和貴, 電気通信大学, 東京都調布市調布ヶ丘 1-5-1, miyazawa@apple.ee.uec.ac.jp



図 1: 統合認知モデルのグラフィカルモデルによる表現

実ロボットによる親子間インタラクションを模した実験を行い、マルチモーダル情報を取得することで学習と評価を行い提 案モデルの有効性を検証する.

## 2. 統合認知モデル

本稿で提案する統合認知モデルを図1に示す.このモデルは, 多層マルチモーダル LDA(mMLDA)による概念形成を中心に 複数のモジュールを統合することで成り立っている [Fadlil13]. 形成した概念を各モジュールが共有することで,複数の認知機 能を統合する.各モジュールは,言語を扱う Hidden Markov Model (HMM),即時的な行動決定を行う強化学習 (REIN-FORCE),行動の時系列をモデル化し長期的な行動計画を行 う Hidden Markov Model (HMM)である.提案するモデル において,単語情報は mMLDA を通じて実世界の情報に接地 している.この単語情報に HMM(Language)で表現された統 語情報を適用することにより,提案モデルはセンサ情報に基づ き文章を生成することができる.また,文章を分解して実世界 の情報を予測することで、ロボットは文章の意味を理解するこ とができる.一方で、行動に関しては、方策に基づいた即時的 な行動と、プランニングに基づいた行動を考える.即時的な行 動では mMLDA が形成した概念を強化学習の状態空間・行動 空間として扱うことで実現される.また、長期的な行動の計画 は HMM(Planning) に表現される時系列情報を用いることで 実現する.次節から各モジュールによる認知機能の実現につい て示し、最後にモデル全体の統合について述べる.

#### 2.1 mMLDA による概念形成

図1の青枠のモジュールが mMLDA を示している.図1に おいて, z は統合概念を表すカテゴリであり,  $z^O$ ,  $z^M$ ,  $z^R$  は それぞれ下位概念に相当する,物体,動作,報酬カテゴリであ る. 上位カテゴリ z は, 下位カテゴリ間の関係性を捉えてい る.これにより、動作、場所、物体など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 [Fadlil13].  $w^O, w^M, w^R, w^{w*}$  は観測データであ り、それぞれ、物体情報、動作情報、報酬情報、言語情報であ る.  $\phi^*$ ,  $\theta^*$  は多項分布のパラメータであり,  $\beta^*$ ,  $\alpha^*$  はディリ クレ分布のパラメータである.物体情報 w<sup>O</sup>は,物体画像から 畳み込みニューラルネットワーク (CNN) により特徴抽出した 4096 次元の情報を用いた.動作情報 w<sup>M</sup> は、ロボットの各動 作を 4 次元の 1-of-k 表現としたものを用いる.報酬情報 w<sup>R</sup> は、ロボットが行動することで得た報酬値を利用し、報酬値r を r < 0, r = 0, r > 0 により場合分けして, 3 次元の 1-of-k 表現としたものを用いる. 言語情報 w<sup>w\*</sup> は、人から与えられ たシーンに対応する文章の Bag of Words 表現である.

#### 2.2 HMM(Language) による言語学習

言語学習は、Attamimi らが提案している mMLDA と HMM の相互学習により行う [Attamimi16]. このモデルにおいて、文 法とは各単語の品詞と語順の情報であり、品詞は概念クラスに よって表現されると考える. 語順はこの概念クラスがどのよ うな順番で並ぶことで文を構成するかを隠れマルコフモデル (HMM) で表現する. この HMM は図 1 の HMM(Language) に示す、遷移確率  $\pi^L$ ,出力確率  $\theta^L$ 、単語  $w_b^w$  の隠れ状態が  $P_b$ である HMM である. 紙面の都合上詳細な説明は省略するが、 HMM(Language) と mMLDA の相互学習により品詞 (概念ク ラスと単語の結び付け)、文法、概念を学習する.

#### 2.3 REINFORCE による方策学習

即時的な行動の学習は強化学習による方策学習を用いて 行う.強化学習には REINFORCE アルゴリズムを用いる [Williams92].図1に示すように、上位概念を強化学習の状 態空間・行動空間として用いる.ある時刻tの上位概念 $z_t$ は 式(1)に示すように観測情報 $w_t^*$ から推定する.

$$z_t \sim P(z_t | w_t^O, w_t^M, w_t^R, w_t^w)$$
(1)

推定した上位概念 z<sub>t</sub> を用いて式 (2) より次の状態の上位概念 z<sub>t+1</sub> を求める.

$$\pi_{\theta_{RL}}(z_{t+1}|z_t) = \frac{exp(\theta_{RL}^T\phi(z_{t+1}, z_t))}{\sum_{z_{t+1} \in A} (exp(\theta_{RL}^T\phi(z_{t+1}, z_t)))}$$
(2)

ここで, A は全行動,  $\theta_{RL}$  は方策  $\pi_{\theta_{RL}}$  のパラメータベクト ル,  $\phi(*)$  は状態  $z_t$  と行動  $z_{t+1}$  のカーネル関数である.  $\theta_{RL}$ の更新は以下の式により行う.

$$\theta_{RL}^{t+1} = \theta_{RL}^t + \eta \nabla_{\theta_{RL}} J(\theta_{RL})$$
(3)

$$\nabla_{\theta_{RL}} J(\theta_{RL}) \approx \sum_{e=1}^{E} \sum_{t=1}^{T} (R_t^e - \bar{b}) \nabla_{\theta_{RL}} \log \pi_{\theta_{RL}} (z_t^e | z_{t-1}^e)$$

(4)

ここで,  $\eta$  は学習率,  $R_t^e$  は e エピソードの t ステップ目の報酬,  $\bar{b}$  は平均報酬である.実際の動作選択や物体選択は,推定した 上位概念  $z_{t+1}$  を用いて  $w^M \sim p(w^M|z_{t+1})$ ,  $z^O \sim p(z^O|z_{t+1})$ としてサンプリングすることで行う.

# 2.4 HMM(Planning) による時系列学習

長期的行動計画の学習は、図 1 上部に示すように、 HMM(Planning)を用いて統合概念 z を時間方向に発展させることで行う.隠れ状態 s,出力確率  $\theta = P(z_t|s_t, \alpha)$ ,遷移確率  $\pi = P(s_t|s_{t-1}, \gamma)$ である HMM を考える.ここで  $\alpha, \gamma$  はそれぞれ、ディリクレ分布のパラメータである.各パラメータ  $\theta, \pi$ をギブスサンプリングによりサンプリングすることで学習を行う.学習を行うことで、時系列情報をモデル化できるため行動計画を行うことができる.この行動計画は、時系列的に欠損された情報を予測することで行う.欠損値は一様分布であると考えてギブスサンプリングすることで長期的な系列を推定し、行動計画が可能になる.

#### 2.5 モデル全体の学習

モデルの結合には中村らが提案している, SERKET のメッ セージパッシングにより行う [Nakamura18]. この枠組みでは 大規模なモデルの更新を各モジュールごとの更新により行い, モジュール間で情報をやり取りすることでモデル全体の学習 を実現する.mMLDA は MLDA の統合により実現する.下 位概念により推定した P(z\*|w<sup>w\*</sup>, w\*) と上位概念により推定 した P(z\*|z) をパッシングし合い相互に更新する. mMLDA と HMM(Language) の統合による言語学習は、ある概念 C における単語 w とカテゴリ k の相互情報量 MI(w, k|C) と, HMM(Language) のパラメータにより重みづけられた単語の 頻度  $P(w^{wC}|w_{h}^{w}, P_{h})$ をパッシングし合い相互に更新すること で実現する. REINFORCE の更新は, MLDA(TOP) により 推論した状態を強化学習のモジュールにパッシングすることで 実現する. HMM(Planning)の更新は MLDA(TOP) により推 定した  $P(z|z^O, z^M, z^R)$  と HMM(Planning) により推定した P(z|s)をパッシングし合い相互に更新することで行う.この ようの各モジュールの更新とメッセージバッシングを繰り返し 行うことで全体の更新を行う. モデル全体を学習することで, モジュール内の予測のみでなくモジュール間での予測が可能に なる.

### 3. 実験

提案手法の有効性を示すために実際のロボットを用いた実験 を行った.実験の目的は、ロボットが環境に対する知識がない 状態から試行錯誤的に概念・行動・言語を学習し、その結果を 解析することでモデルの妥当性を検証することである.

#### 3.1 実験設定

図2に実験設定を示す.実験には双腕ロボットBaxterを用いた.ロボットは図2(a)に示す物体に対して,図2(b)に示す行動から図2(c)に示す報酬に沿った適切な行動を選択するように学習する.物体は図2(a)に示すように,3つのグループに分けた.学習には学習データ1及び2を,評価には評価データを用いた.図2(b)に示す動作は人手により事前に設計した.

図3に実験の流れと概要を示す.ロボットは机上から物体 を一つ選択し,適切な動作を選択するように学習する.物体を



	ぬいぐるみ	ボール	マラカス	スプレー缶		
掴む	0	0	0	0		
片付け	1	1				
にぎる	1 1		-1	-1		
振る	3 -1 -1 1 I					
*物体把持時に「掴む」を選択,続けて同じ行動→報酬値-1 (c)物体と動作に対応した報酬値の設定						
図 2: 実験環境の設定						

箱に片付けると机上の物体が一つ減る.すべての物体を片付 けるともう一方の学習データの物体セットが机上に置かれる. ロボットが行動すると,親役の実験者は物体と動作に対応した 言語情報と報酬情報を与えた.この環境では物体を把持し,適 切な遊び(「にぎる」または「振る」)を行い,箱に片付ける という行動を繰り返すと最も高い報酬値が得られる.

次に,行動決定と情報取得の詳細について説明する.ロボットは図3に示すように行動決定と情報の取得を行う.ロボットの行動決定は物体を把持しているか否かにより,物体選択と動作選択の2通りに分かれる.物体を把持していない場合には,物体選択を行う.机上の各物体の物体概念 *z*<sub>ID</sub> と方策により決定した上位概念 *z* から求めた *P*(*z*<sup>0</sup>|*z*) のコサイン



表 1	:下位	概念の	カテ	ーゴリ	り分類	精度
-----	-----	-----	----	-----	-----	----

	物体概念	動作概念	報酬概念
学習データ 1,2	0.81	1.00	1.00
評価データ	0.68	1.00	1.00



図 4: 各エピソードごとの累積報酬値

類似度を計算する.最もコサイン類似度が高い物体を把持する.ロボットが物体を把持している場合には,動作選択を行う.動作は,方策により求めた上位概念 z から動作情報  $w^M$  を  $w^M \sim P(w^M|z) = \sum_{z^M} P(w^M|z^M) P(z^M|z)$  とすること で選択する.ロボットはすでに物体を把持しているため,物体情報  $w^O$  は 1 ステップ前の物体情報  $w^{O_1}_{t-1}$  とした.

ロボットが上記の行動選択を行うと、教示者はロボットが行った動作と物体に対応した言語情報と報酬情報を与える.ユーザーは、ロボットがボールを握って正の報酬を得た場合には、「ボールを握るのは良いね、それは柔らかいボールだよ」などのような発話を行った.この発話に対して、音声認識による文章取得と MeCab による形態素解析を行うことで、 $w^w$ を取得する.報酬情報は、教示者が図 2(c) に示す報酬値を直接与えた.これにより $w^w$ ,  $w^R$  を取得する.上記の行動により、ロボットはマルチモーダル情報 $w^O$ ,  $w^M$ ,  $w^R$ ,  $w^w$  を取得する.取得した観測情報と方策 $\pi_{\theta_{RL}}$ に基づいて、次状態の上位概念 $z_{t+1}$ を求める.これは、式(1)、式(2) に基づいて行った.

上記の行動決定から次状態の行動決定までの流れを1ステッ プとして実験を行った.実験は1エピソード50ステップとし, 10エピソード,計500ステップを行った.1エピソードごとに 取得したマルチモーダル情報 w\*を用いて統合認知モデルのパ ラメータを更新した.これを繰り返すことでロボットは概念・ 行動・言語を学習する.また,11エピソード目に評価データ を用いて行動選択を行い,モデルの汎化性を検証した.

# 4. 実験結果

概念形成と行動学習,言語学習についての実験結果を示す. その後モジュールの統合による影響について述べる.行動計画 に関する結果は紙面の都合上省略する.

### 4.1 概念形成と行動学習

まず,概念のカテゴリ分類精度について表1に示す.学習 データと評価データに対する分類精度を各下位概念ごとに示し ている.これを見ると入力情報が単純な物体・報酬概念につい ては全て分類が正しい.物体概念は分類誤りが含まれるがチャ ンスレベルである0.25と比較すると高い値となっている.次 に行動学習につてい図4に示す.これは各エピソードごとの累 積報酬値を示している.この図を見ると,エピソードを得るご とに累積報酬値が上昇していることが分かる.また,評価デー タに対しても学習データと同様の累積報酬値を得ていることか ら,汎化性のある行動選択が獲得できたといえる.

表 2: 文章生成における評価値

	又法	物体	」 動作	報酬
正しい文章数	48	24	43	50
間違えの文章数	2	26	7	0

表 3: 言語情報から実世界情報の予測精度

	物体概念	動作概念	報酬概念
モデル1	1.00	1.00	1.00
モデル3	0.83	0.54	1.00

#### 4.2 言語学習

言語理解と文章生成のタスクを行い言語学習の評価を行った. まず,言語理解について示す. 文章情報 w<sup>w</sup> から w<sup>w</sup> が接地し た実世界情報 w<sup>O</sup>,w<sup>M</sup>,w<sup>R</sup> を予測するタスクを行い言語理解に ついての評価を行った.「ぬいぐるみを振るのはいいね」という ような文章情報を 24 文用意し、それを入力として  $w^O, w^M, w^R$ の予測を行い、文章が指し示す情報の予測できるかを検証し た. その結果, 24 文章すべての文章に対して w<sup>O</sup>,w<sup>M</sup>,w<sup>R</sup> を 正しく予測できた.次に文章生成について示す.これは実世界 情報  $w^O, w^M, w^R$  から文章情報  $w^w$  を予測することで行った. 評価データ 50 ステップの実世界情報  $w^O, w^M, w^R$  を入力とし てその情報を説明するような文章が生成できるかを検証した. 各生成した文章に対して, 文法的に正しいか, 物体・動作・報 酬情報それぞれについて正しく説明できているかを評価した. その結果を表2に示す.50文章中ほとんどの文章は物体情報 を除いて正しい説明ができている.物体情報に関して正解とな る文章の生成数が少なかったのは,物体概念の分類精度が低い ことに起因する物体概念に結びついた言語情報の予測性の低さ が問題であると考える.

### 4.3 モジュールの統合による影響

モジュールの統合による影響を検証するために4種のモデル を比較した.ぞれぞれ、モジュール欠損なし、HMM(Language) が欠損,HMM(Planning)が欠損,HMM(Language)と HMM(Planning)が共に欠損したモデルである.各モデルの 学習時の対数尤度の変化について図5に示す.これを見ると、 下位概念の対数尤度が言語モジュールを含むモデル1,2では上 昇していることがわかる.また、各HMMのモジュールは更 新回数が進むことで対数尤度が上昇していることがわかる.

言 語 理 解 に つ い て 予 測 性 を 評 価 す る タ ス ク を HMM(Language) が 欠損して いるモ デル 3 につ いても 行い予測性の違いについて検証した.その結果を,表 3 に示 す.この結果を見ると前述したようにモ デル 1 はすべての文 章に対して正しい予測が行えている.モデル 3 では物体,動 作について予測精度の低下がみられる.つまり,言語モジュー ルを統合することで予測性能が向上した.これは,mMDLA に入力される言語情報 w<sup>w\*</sup> が HMM(Language) により正し く重みづけされたため予測性が向上したと考えられる.

### 5. 結論

本研究では、概念・行動・言語を学習するための統合認知モ デルを複数の確率モデルを統合することで実現した.実ロボッ トによる実験を行い、マルチモーダル情報を取得することで学 習と評価を行い提案モデルの有効性を検証した.結果として、 ロボットは自身の経験に基づいて概念・行動・言語を学習した. また、モジュールの統合により言語を含んだ予測タスクの性能 が向上することを確認した.今度は、他者を考慮した文章の生 成や推論などのより高次な認知活動の実現を目指す.



図 5: 各モジュールの対数尤度の変化. 横軸はモデル全体の更 新回数. 縦軸は各モジュールの対数尤度.

#### 謝辞

本研究は JST CREST (JP-MJCR15E3),新学術領域「認 知的インタラクションデザイン学」(26118001)の支援を受け たものです.

### 参考文献

- [Arandjelovi17] R. Arandjelovi, A. Zisserman, "Look, Listen and Learn," arXiv 1705.08168, 2017.
- [Attamimi16] M. Attamimi, et al., "Learning Word Meanings and Grammar for Verbalization of Daily Life Activities Using Multilayered Multimodal Latent Dirichlet Allocation and Bayesian Hidden Markov Models," Advanced Robotics 30(11-12), pp.806–824, 2016.
- [Fadlil13] M. Fadlil, et al., "Integrated Concept of Objects and Human Motions Based on Multi-layered Multimodal LDA," in proc. of IROS 2013, pp. 2256–2263, 2013.
- [Gu16] S. Gu, et al., "Deep Reinforcement Learning for Robotic Manipulation," arXiv 1610.00633, 2016.
- [Hill17] F. Hill, et al., "Understanding Grounded Language Learning Agents," arXiv 1710.09867, 2017.
- [Levine16] S. Levine, et al., "End-to-End Training of Deep Visuomotor Policies," Journal of Machine Learning Research, Vol.17, pp.1-40, 2016.
- [Miyazawa17] K. Miyazawa, et al., "Integration of Multimodal Categorization and Reinforcement Learning for Robot Decision-Making," IROS2017: Workshop on ML-HLCR, 2017.
- [Nakamura09] T. Nakamura, et al., "Grounding of Word Meanings in Multimodal Concepts Using LDA," IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3943-3948, 2009.
- [Nakamura18] T.Nakamura, *et al.*, "Serket: An architecture for connecting stochastic models to realize a large-scale cognitive model." Frontiers in neurorobotics, 2018.
- [Taniguchi16] T. Taniguchi, et al., "Symbol Emergence in Robotics: A Survey," Advanced Robotics, Vol.30, 11-12, pp.706-728, 2016.
- [Williams92] R. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine Learning*, Vol. 8, Issue 3, pp. 229-256, 1992.