



Title	好奇心の赴くままに行動して学習するロボット
Author(s)	
Citation	令和4（2022）年度学部学生による自主研究奨励事業 研究成果報告書．2023
Version Type	VoR
URL	<a href="https://hdl.handle.net/11094/90993">https://hdl.handle.net/11094/90993</a>
rights	
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

## 令和4年度大阪大学未来基金「学部学生による自主研究奨励事業」研究成果報告書

ふりがな 氏名	やなぎだ かんご 柳田 栞吾	学部 学科	基礎工学部 システム科学科	学年	3年
ふりがな 共同 研究者氏名		学部 学科		学年	年
アドバイザー教員 氏名	堀井 隆斗	所属	基礎工学研究科		
研究課題名	好奇心の赴くままに行動して学習するロボット				
研究成果の概要	研究目的、研究計画、研究方法、研究経過、研究成果等について記述すること。必要に応じて用紙を追加してもよい。(先行する研究を引用する場合は、「阪大生のためのアカデミックライティング入門」に従い、盗作剽窃にならないように引用部分を明示し文末に参考文献リストをつけること。)				

## 1. はじめに

近年、人と隔絶された生産工場内だけでなく、人が生活する家庭環境内でもロボットを活用しようと研究・開発が盛んに行われている。例えば、自律ロボットによるサービス支援技術の向上を目的とした RoboCup@Home [1]では、部屋の片付けや人が指定した物を取ってくる課題などの達成度を競う。近年の深層学習による画像処理の成功に伴って、高い精度で物体を認識することができるようになり、部屋の片付け課題をほぼ完了させるチームも出てきている。しかし、現在のロボットが複雑な課題をこなすためには、課題達成の条件を設定するだけでなく、我々にとっては常識だと思われる知識を設計者が作り込まなければならない。このような知識の例として、例えば物体の呼び方がある。日常生活において、我々は物体をその名前と呼称することもあれば、その特徴や性質で呼称することもある。「おもちゃ置き場から“もふもふしたもの”を取ってきて」と言われたとき、人であれば自らの知覚経験を通じてぬいぐるみを指していると理解できるが、物体の名前だけを知識として埋め込まれたロボットにはそれができず、また、様々な物体と感覚の関連付けを全て事前に設計しておくことは不可能である。

では、我々の知識はいかにして獲得されるのであろうか。人の学習や発達過程において、知識は他者から与えられるものだけではなく、自ら学ぶものであることがわかっている。特に、乳幼児は我々がつきっきりで知識を与えずとも自らの好奇心に従ってさまざまな情報を収集し学習していくことが知られている[2]。

そこで本研究では、人の発達過程や、世界を知りたいという普遍的な欲求である好奇心のモデル化を通じて、ロボットの知能化を試みた。具体的には、まず、自ら世界に対して行動し、感覚情報を得るために必要な身体であるロボットアームを設計製作した。次に、物体に対する知覚-運動経験から似

た経験を集めて、重要な特徴で抽象化したものである物体概念を形成するための計算モデルを選定し、作成したロボットによって収集したデータをもとに形成した概念構造を解析する実験を行った。さらに、乳児が試行錯誤的な運動を通じて身体を制御する方法を学習するのに重要な現象である身体バブリングや、無意識に環境に対して反射行動をとる原始反射を強化学習によってモデル化し、好奇心を満たす、すなわち世界を知るのに必要な行動の獲得と概念形成を通じた世界の構造的理解を相補的に行う実験をシミュレーションで簡易的に行った。

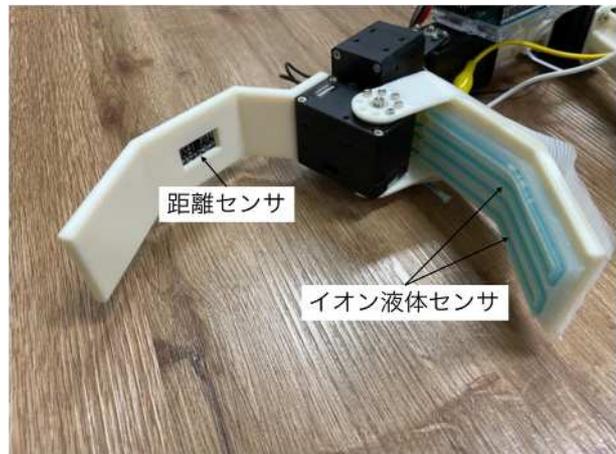
## 2. 自ら感覚情報を得て学ぶロボットアームの設計製作

本研究ではまず、マルチモーダル知覚-運動経験を自らの知識として学習するためのプラットフォームとして、様々なセンサを有するロボットアームを開発する。市販のロボットアームでは、身体バブリングを実現するための関節可動域が狭く、手頃な大きさのものが無いため、オープンソースの小型ロボットである ROBOTIS OpenMANIPULATOR-X を参考に、ハンドの関節を合わせて 6 自由度を持つロボットアームを設計製作した。図 1 にその外観を示す。本実験でロボットは床に置いた物体を掴む想定であることから高さ方向に対する動作は不要であるため、スカラ型ロボットアームにした。これにより物体を掴む動作を効率的に学習することができ、さらに奥行方向への作業可能領域を拡張できる。また、ハンドは固定指と可動指からなる単純なものにすることで、可動指を開いた状態で物体に近づいて押しつければ、自動的に物体がハンド内に収まるため、物体を掴むまでの動作が容易になった。

ハンドにはそれぞれの指に物体情報を取得するためのセンサを配置している。可動指には、物体概念の獲得に重要となる触覚情報を得るためのイオン液体センサ[3]を配置している。ここでは、細田研究室の川節助教の協力で、現在研究されているイオン液体センサを使用した。このセンサは、柔らかい材質でできた経路をイオン液体で満たしたもので、経路端の電極に電圧を印加することでイオン液体を抵抗とした回路ができる。外力によって経路が狭まると抵抗値が上昇する性質を利用することで、物体の硬さなどの触覚情報を得ることができる。しかしイオン液体センサでは、物体が触れたことを検知できないため、固定指に近距離を測定する距離センサを埋め込んだ。物体との距離が、ある閾値以下になれば固定指に物体が触れたとする。なお、距離センサによる感覚を以降では近接覚と呼ぶ。また、視覚情報を得るためのカメラはハンドと向かい合う形でロボットとは独立して配置した。



(a) 全体



(b) ハンド部

図 1 設計製作したロボットアーム

### 3. マルチモーダル概念形成

人は、物体に対してある行動を取った結果から得られるマルチモーダルな感覚情報を統合し、それらをカテゴライズすることで概念を形成する。また、この概念に基づいて、未知物体に対する情報を予測することができる。つまり1つの答えとして知能とは、概念形成を通じて世界を理解し、予測し続けることであるといえる。本研究では、このような概念形成のための計算モデルとして multimodal latent Dirichlet allocation (MLDA) [4]を用いる。MLDA のグラフィカルモデルを図2に示す。 $z$ は概念カテゴリであり、 $\theta$ をパラメータとする多項分布から生成される。 $w^v, w^q, w^t$ はそれぞれ、視覚、作成したロボットアームのハンドの関節角、触覚センサの観測情報であり、それぞれが $\phi^v, \phi^q, \phi^t$ をパラメータとする多項分布から生成されると仮定する。 $\alpha, \beta^*$ は多項分布の共役事前分布であるディリクレ分布のパラメータである。視覚情報 $w^v$ は、画像から畳み込みニューラルネットワーク (convolutional neural network: CNN) で抽出した特徴量を用いる。関節角情報 $w^q$ 、触覚情報 $w^t$ は vector quantized variational autoencoder (VQ-VAE) [5]により得られる表現を用いる。VQ-VAE はエンコーダで抽出した特徴量をコードブックで離散表現にすることで、デコーダでの再構成性能を VAE より向上したモデルである。本研究では、1次元 CNN をエンコーダ、デコーダとする VQ-VAE を使用して、関節角と触覚センサの時系列データから得られる離散表現の頻度情報を $w^q, w^t$ として用いる。

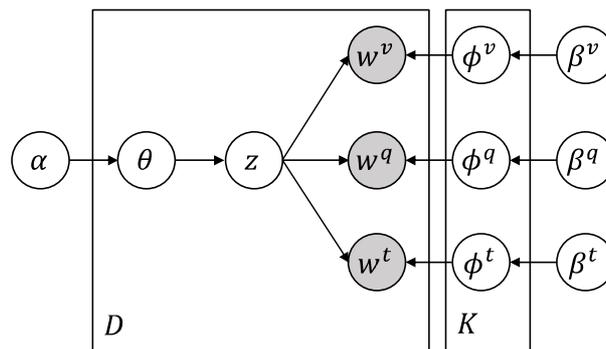


図2 MLDA のグラフィカルモデル

### 4. 強化学習による行動の学習

人の発達初期に見られる現象として、身体全体をランダムに動かす身体バブリングや、環境に対して無意識に反射行動をとる原始反射が挙げられる。これらの現象に加えて、乳幼児は世界を知りたいという好奇心によって、外界に対して行動する能力、すなわち感覚運動連関を学習し、感覚情報の変化量が大きくなるような行動を自律的に獲得することができる。ここでは、強化学習を用いてこれらをモデル化する。強化学習は、ある一定時間に行動した結果、得られると期待される報酬の総和、すなわち期待収益を最大化するような行動を決定する方策を学習する問題である。この報酬を好奇心と対応づけて、内部報酬とすることで、ロボットが自律的に動作を獲得することを目指す。具体的には、ロボットは、目の前の物体についてよく知りたいという好奇心に基づいて、物体に手を伸ばし（リーチング）で掴み（グラスピング）、触覚情報と視覚情報を得ようとする。これらを継続的に行うことで、ロボットはリーチング動作とグラスピング動作を獲得し、さらに MLDA による物体の概念形成を通じて世界を理解していく。本研究では、強化学習の学習手法として Soft Actor-Critic (SAC) [6]を用いる。SAC は期待収益と方策エントロピーの $\alpha$ 倍の和を最大化する。ここで $\alpha$ は方策エントロピー係

数で、学習が進んで期待収益が大きくなるにつれて小さくなっていく。学習初期は方策エントロピーが大きいため、行動は確率的、すなわちランダムに行動し、感覚と運動、報酬の関係を学習する。学習が進むと、方策エントロピーは小さくなっていき、報酬が大きくなる行動を決定的に取る。これは乳児が身体バブリングによって動作を獲得し、自分の意思で行動する過程に相当する。さらに、学習初期に物体がハンドに触れた（近接覚センサによって判断）ときには、反射的にハンドを閉じる動作をロボットに組み込む。これは、乳児のグラスピング動作の獲得に重要な原始反射である手掌把握反射に相当する。

## 5. 実験

ここでは、まず実環境での予備実験として、あらかじめ設定したグラスピング動作によって得られたマルチモーダルな感覚情報を用いた MLDA による概念形成の実験について述べる。次に強化学習による行動の学習と MLDA による概念形成を相補的に行う実験について説明する。ただし、後半の実験については実施時間の関係上、実験設定を簡略化したものをシミュレーション上で実施したので、その結果を報告する。

### 5.1. 予備実験：MLDA を用いたロボットアームによる概念形成

#### 5.1.1. 実験の目的

本予備実験の目的は、作成したロボットアームが収集可能なデータで MLDA が学習できるかを確認することである。ハンド内に物体がある状態、つまりリーチング動作まで完了した状態で、あらかじめ設定したグラスピング動作をロボットにさせて、複数の触感や色を持つボールのマルチモーダルデータを収集する。収集したデータを用いて MLDA を学習させた後にその概念構造を解析する。

#### 5.1.2. マルチモーダルデータの収集

マルチモーダルデータの収集に使用した物体を図 3 に示す。ここでは、材質の異なる 4 種類のボールとして、左上のプールボール、右上のマッサージ用ボール、左下の犬用ぬいぐるみボール、右下のスポンジボールを用意した。以降それぞれ、pool, massage, dog, sponge と呼ぶ。なお、dog を除くその他のボールはほぼ同じ大きさのものを用意した。

4 種類の材質と色違いを含めて全 23 個のボールについて、それぞれ 5 回ずつ、計 115 回のグラスピング動作時のマルチモーダルデータを収集した。図 4 にデータ収集の環境を示す。まず、左のカメラで物体の画像を撮影した。次に物体を 1 秒間握り 1 秒間離す動作を 5 回、すなわち 10 秒に渡ってグラスピングしたときのハンド関節角と触覚情報を 20 Hz で収集した。収集した物体画像の例を図 5、ボールの各種類で平均したハンド関節角と触覚情報の時系列データを図 6 に示す。関節角の値が小さいほどハンドを閉じている。一方で触覚情報は、物体が硬いほど大きな値をとる。

#### 5.1.3. モデルの設定

本実験では、MLDA のディリクレ分布のパラメータ  $\alpha, \beta^*$  を全て 1.0、 $z$  の持つカテゴリ数  $K$  を 4 種類とした。データ数  $D$  は 115 個である。画像から視覚情報  $w^v$  を抽出するための CNN には学習済みの GoogLeNet [7] を使用した。5.1.2 節で収集したマルチモーダルデータから関節角情報  $w^a$  や触覚情報  $w^t$  を抽出するための VQ-VAE は、あらかじめランダムな三角関数 10 個を重ね合わせた時系列データ 5 万個で事前学習したものを使用した。視覚情報  $w^v$  には GoogLeNet によって抽出した 1024 次元の特徴ベクトル、関節角情報  $w^a$ 、触覚情報  $w^t$  には、128 のコードブックを持つ VQ-VAE によって得られた離散表現の頻度情報を使用した。



図 3 実験に使用した物体

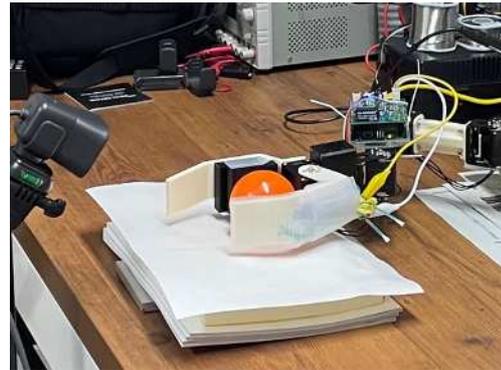


図 4 データ収集の環境



(a) 赤色の dog



(b) 橙色の massage

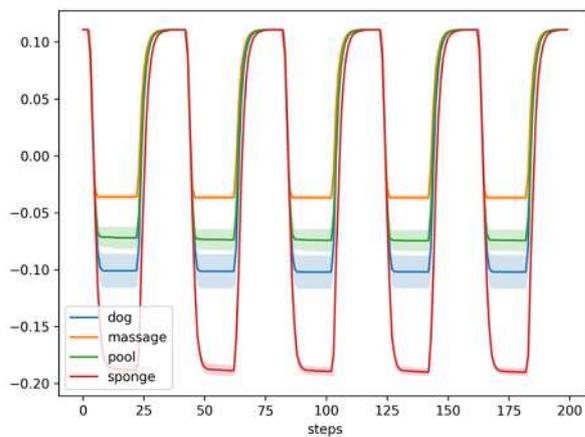


(c) 緑色の pool

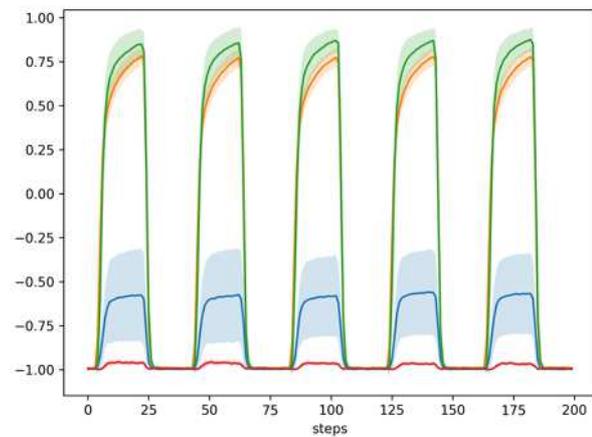


(d) 青色の sponge

図 5 ボールの画像例



(a) ハンドの関節角



(b) 触覚

図 6 各感覚情報の平均時系列データ

### 5.1.4. 結果

MLDA を学習させる条件として、①視覚のみ、②視覚と関節角、③視覚と触覚、④関節角と触覚、そして、⑤全ての感覚情報を用いた場合の5つを用意した。それぞれ異なるシードで5回学習させた結果のカテゴリ一致率の平均を図7に示す。また、最もカテゴリ一致率の高かったシードでの学習によって形成した概念をPCAで2次元に圧縮したものを概念空間として図8に示す。2つ以上のモダリティを用いた場合は、単一モダリティである視覚のみの場合よりも一致率が高かった。これは、モダリティが増えることで、カテゴリ固有の情報が増え、より一層普遍的に物体概念が形成されるからである。実際に、図8(a)の視覚のみで形成された概念空間において、青丸上で概念が形成された物体を図9に示すが、物体の色である黄色が概念の形成に影響を強く与えており、色に対する普遍性が無いが、モダリティが増えた場合の概念空間(図8(b)~(e))では、同じカテゴリで大きく飛び越えて形成されているものは少なくなっており、色に対する普遍性があると考えられる。一方で、標準偏差が最も小さかったのは、視覚のみの場合である。これは、一般的にCNNは空間的なばらつきに対してロバストであるからだと考えられる。また最も標準偏差が大きいのは関節角と触覚の場合である。これは、図6に示す関節角と触覚データにばらつき、特に触覚ではmessageとpoolに重なりがあるデータが多いからである。最も一致率が高かった全感覚の場合の概念空間(図8(e))は、カテゴリごとに分布しており、物体概念の形成がよくできている。図8(b)を見ると、横や縦軸の斜め45度の軸がハンドをどこまで閉じたかの情報、つまり物体の柔らかさに対応している傾向にある。また図8(c)~(e)を見ると、縦軸が相対的な物体の柔らかさに対応している。ただしあくまでも、横軸に関してほとんど同じ位置のクラス同士、すなわちdogとsponge、messageとpoolに対して相対的にどちらが柔らかいかを表している。逆に横軸は大まかな物体の柔らかさを表しているといえる。なお、図8(c)と図8(d)、(e)でmessageとpoolの関係性が上下逆になっている理由は、関節角で見ればpoolの方が柔らかいが、触覚で見ればmessageの方が柔らかいデータになっていたからである。関節角データの方が両者に差があるため、図8(d)、(e)ではその情報が強く反映されている。さらに図8(c)、(f)の横軸は、先ほど述べたように大まかな物体の柔らかさに対応しているだけでなく、物体表面の光沢度に対応しているともいえる。今回使用した物体では表面に光沢のあるmessageとpoolの方が光沢のないdogとspongeに比べて非常に硬いため、光沢度の高い物体は硬いという概念が形成されている可能性が高い。

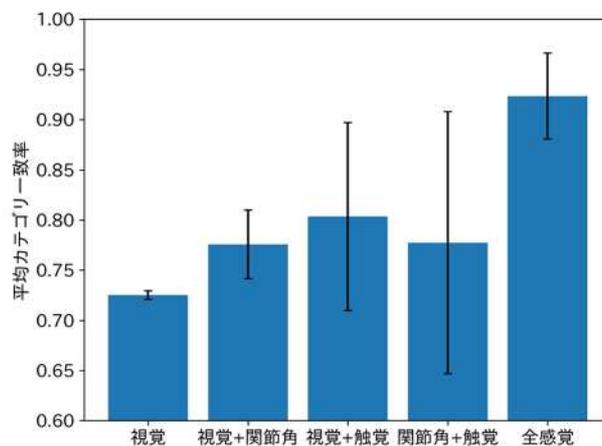


図7 各条件における平均カテゴリ一致率

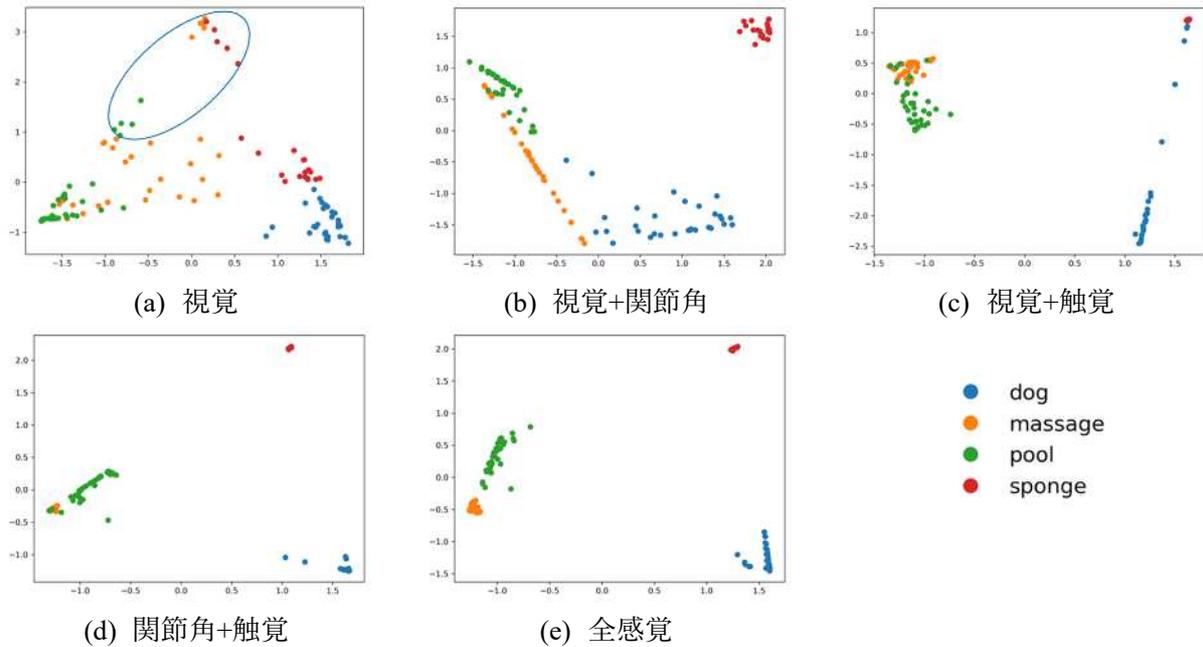


図 8 PCA で 2次元に圧縮した概念空間

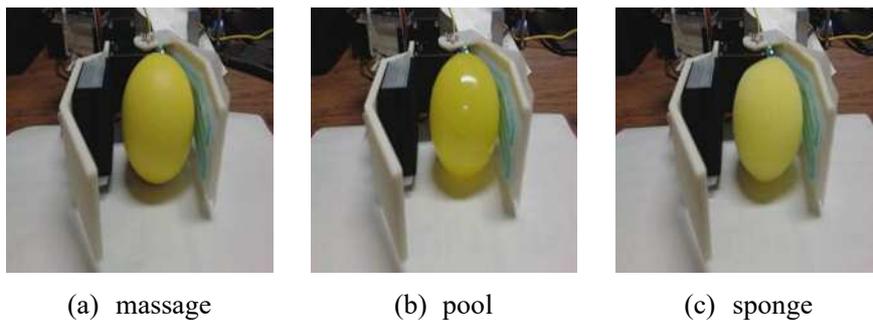


図 9 カテゴリの異なる黄色のボール. GoogLeNet に入力するサイズ(224×224)に変更.

## 5.2. 強化学習による行動獲得と概念形成の相補的学習

### 5.2.1. 実験の目的

本実験の目的は、強化学習を用いて、世界を知るのに必要な行動の獲得と MLDA による概念形成を通じた世界の構造的理解を相補的に行うことで、ロボットが好奇心の赴くままに行動して学習することが可能かどうか検証することである。強化学習では試行錯誤的に動かすことで行動を学習するため、実環境では、ロボットが壊れる危険性が高い。さらにセンサの値に不確実性があることから、本実験はシミュレーション環境で行う。

### 5.2.2. シミュレーション環境

作成したロボットアームの 3D モデルを強化学習のシミュレータとしてよく使用されている MuJoCo [8] の環境で使用した。図 10 に作成した MuJoCo 環境を示す。ロボットの周りには柵を設けており、配置した物体がロボットアームの作業可能領域外に出ないようにした。また、図 3 中の物体の関節角や触覚特性を模した柔軟物体を 4 種類作成した。

### 5.2.3. 学習設定

本実験の課題は、作業領域内にランダムに置かれた物体にリーチングし、グラスピングして得られ

た感覚情報をもとに、物体がどのカテゴリに属しているかを予測することである。ロボットは感覚情報として、6つの関節角度と手先位置、物体位置、近接覚、触覚の情報を得る。また、行動（ロボットの運動指令）は6つの関節の目標関節角度とし、0.05秒毎に感覚情報を得て次の行動を決定するものとした。ただし、学習初期にリーチングに成功したときは、手掌把握反射として物体をグラスピングし続ける行動が割り込みとして発生するようにした。学習を効率化させるための報酬として、リーチング動作に対して、手先位置と物体位置の差の負のL2ノルム、グラスピング動作に対して、何らかの触覚情報を得られれば+1、得られなければ0と設定した。さらに好奇心に対応する内部報酬には、物体が属するカテゴリの予測確率の負の情報エントロピーを設定した。なお、グラスピング動作までできなかった場合は、カテゴリの予測確率を全て同じ確率として、負の情報エントロピーを計算した。つまり本実験では物体のカテゴリ数は4なので、グラスピングできない場合の負の情報エントロピーは、 $\sum_{i=1}^4 0.25 \ln 0.25 = \ln 0.25 \approx -1.39$ になる。

1エピソードにつき、1つの物体を作業領域内に配置し、タスク中20秒間のうち10秒間物体をグラスピングできた場合は、そのときの触覚情報をMLDAの学習データとして保存した。なお視覚情報は、5.1節で収集した画像の中から、そのエピソードで配置した物体カテゴリに対応するものをランダムに選択した。また、MLDAやCNN、VQ-VAEは5.1節で使用したものと同一ものを利用した。MLDAは、4カテゴリの物体を順番に配置するタスクを2回繰り返した後、すなわち8エピソード毎に5.1節と同様に5回学習した結果を平均した。

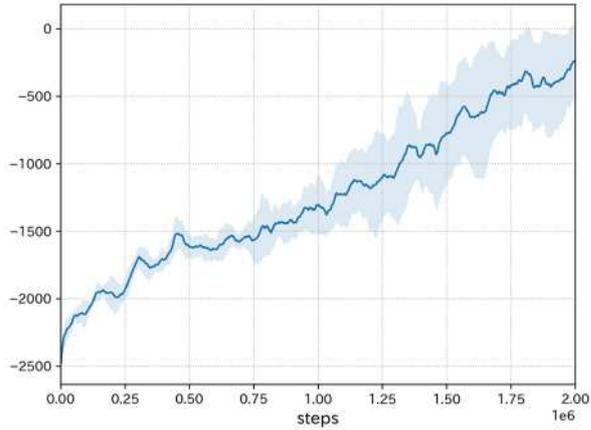


図10 作成したMuJoCo環境

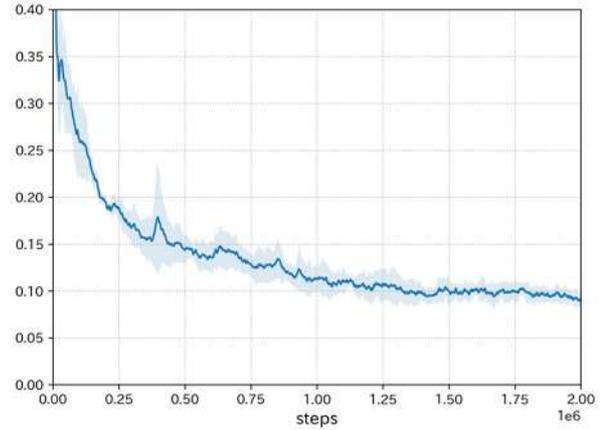
#### 5.2.4. 結果

異なるシードで4回、200万ステップ学習させた結果の平均を図11に示す。図11(a)より報酬は200万ステップ以降も上昇すると考えられる。これはリーチングに失敗するエピソードがまだあることを意味している。図11(b),(c)より40万ステップまでは、方策エントロピー係数の急激な減少とデータ収集の失敗が確認できる。これは40万ステップの間に身体バブリングを中心に行うことで、一部のリーチング動作を獲得し、それ以降のステップで、物体にリーチングした後に手掌把握反射によってグラスピングまで成功して、触覚情報を得るようになってきていることを意味している。図11(d),(e)より、80万ステップ頃には平均カテゴリ一致率が90%程度になり最終的に96,97%程度で頭打ちしているが、標準偏差は減少し続けている。また、図11(f)より、物体の属するカテゴリの予測確率の情報エントロピーの平均値が減少し続けていることがわかる。これは、物体のグラスピングに失敗したエピソード数が減少していることにも関わるが、情報エントロピーがほとんど0になっているエピソードも存在することから、高い精度でグラスピングした物体のカテゴリを予測できていることを意味している。学習を重ねるにつれて概念が形成される過程の一例を図12に示す。なお図の説明は、その時の

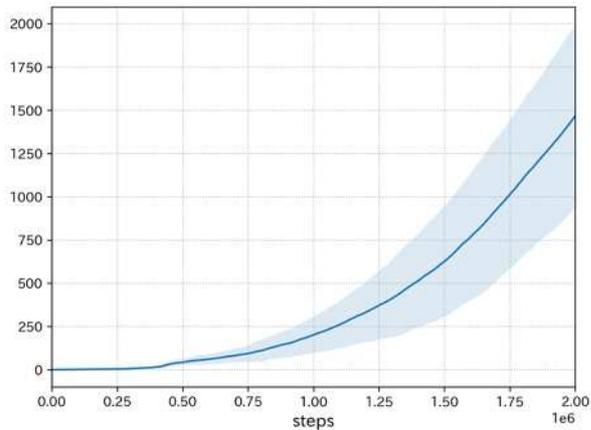
学習したステップ数と取得した物体のデータ数を意味している。学習初期 (図 12(a)~(c)) からカテゴリごとに分布しているが、最も柔らかさの異なる *massage* と *sponge* が近い概念として形成されており、良い概念空間ではない。学習中盤以降 (図 12(d)~(f)) では、物体に対する知覚-運動経験が増えることで、良い概念が形成されていることがわかる。最終的に獲得された概念空間 (図 12(f)) を見ると、横軸から斜め 45 度の軸が相対的な物体の柔らかさに、それに直交する軸が大まかな物体の柔らかさに対応しており、5.1 節の結果と同様の傾向があることが確認できる。



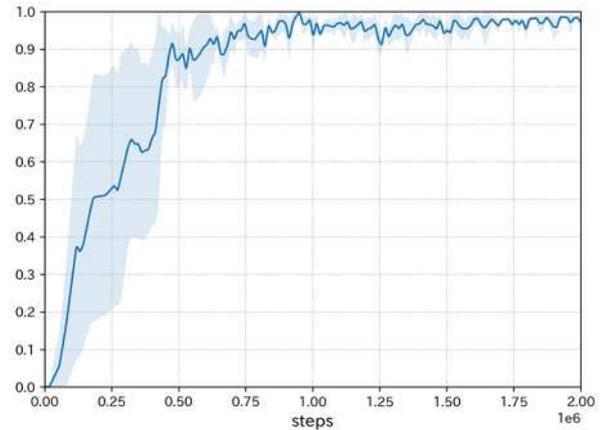
(a) 報酬



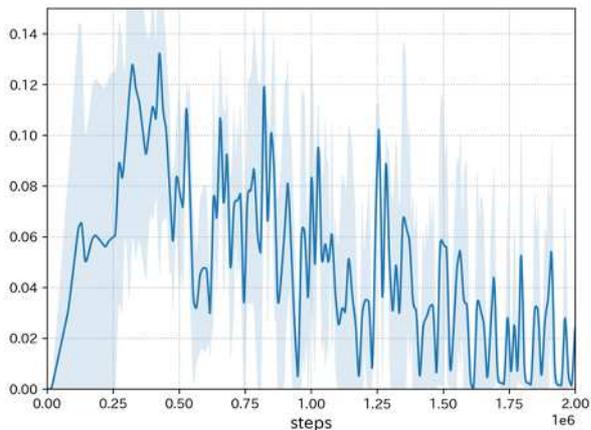
(b) 方策エントロピー係数



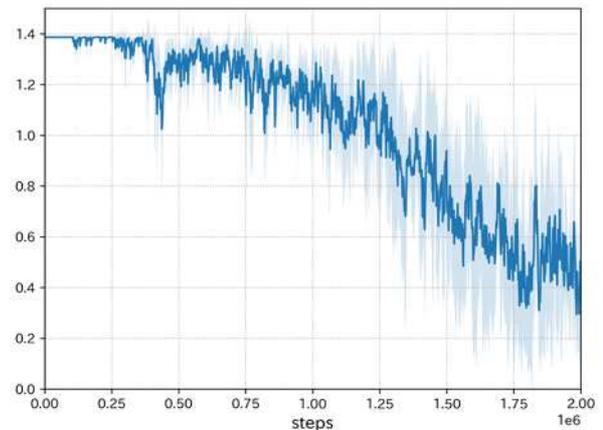
(c) 取得した物体のデータ数



(d) 平均カテゴリ一致率



(e) カテゴリ一致率の標準偏差



(f) 情報エントロピー

図 11 異なるシードで 4 回学習させた結果の平均

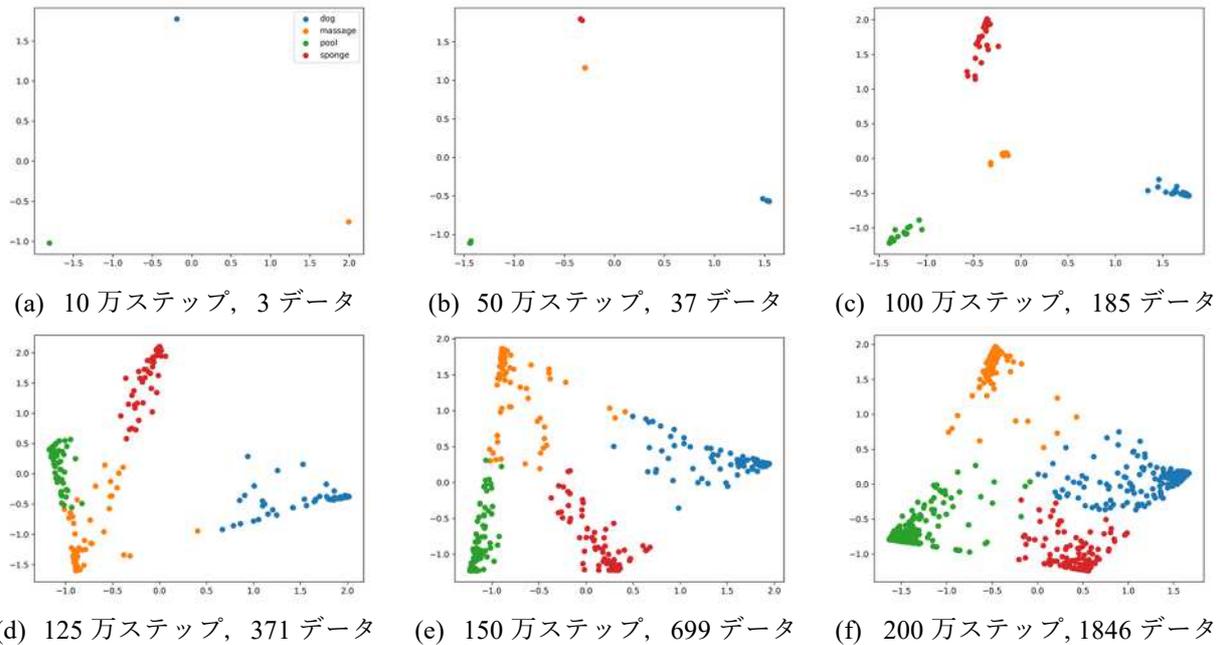


図 12 概念が形成される過程の一例

## 6. まとめと今後の課題

本研究では、人から知識を与えるのではなく、自ら行動して得た情報をもとに学習するロボットを実現するために、乳児の発達過程や、世界を知りたいという普遍的な欲求である好奇心の計算モデルを作成したロボットアームに組み込み、① MLDA を用いたロボットアームによる概念形成の実験と、② 強化学習による行動の学習と概念形成を通じた世界の構造的理解を相補的に行う実験、の 2 つを行った。

① MLDA を用いたロボットアームによる概念形成の実験では、作成したロボットアームで、あらかじめ設定したグラスピング動作によって収集したマルチモーダルデータを使用して MLDA の学習を行った。モダリティを増やすことで、視覚のみでは概念形成に影響を受けた物体の色は、不必要な特徴として捨象され、より重要な特徴で抽象化されたため一致率が向上した。また複数モダリティで形成された概念空間では、物体の柔らかさに対応する軸が存在した。特に視覚と触覚情報を含んだ場合は、表面の光沢度が高い物体は硬いという概念が形成されている可能性が高いことが確認できた。これは、我々が持っている、一般的に光沢のある物体は硬い傾向にあるという経験をロボットも獲得したといえる。

② 強化学習による行動の学習と概念形成を通じた世界の構造的理解を相補的に行う実験では、ロボットが好奇心の赴くままに行動して学習することが可能かどうか、シミュレーションで簡略的に検証した。学習初期に身体バブリングを中心に行うことで、リーチング動作を獲得し、さらに手掌把握反射によってグラスピングまで成功して、触覚情報を得るようになった。物体に対する知覚-運動経験が増えるにつれて、より良い概念が形成され、高い精度でグラスピングした物体のカテゴリを予測できた。しかし、好奇心に対応する内部報酬として、物体が属するカテゴリの予測確率の負の情報エントロピーを設定したが、同時に、学習を効率化するために負の L2 ノルムや触覚情報に対する報酬も設定しており、これは好奇心といった内発的な動機ではなく、外発的な動機であるといえる。そのため、ロボットが自発的に行動して学習したと断言はできない。推測ではあるが、内部報酬がなくても

MLDA による物体概念の形成はうまくいくであろうと考えられる。

以上より今後の課題は、学習を効率化するために設定した報酬をなくして、その代わりとなるような内部報酬を設定することである。具体的には、ロボットの身体に関する内部モデルとして、現在の感覚情報と運動から次時刻の感覚情報を予測する予測モデルを用意し、方策と同時に学習させる。予測モデルは予測値と観測値の差である予測誤差を最小にするように学習するが、この予測誤差を内部報酬とすることで、ロボットは予測が難しい行動を積極的にとり、自身の身体を含んだ世界を構造的に理解していくと考えている。さらに、シミュレーション上で学習した方策を実世界に転移する sim2real の技術を用いることで、実環境で行動を学習する時間の短縮やロボットの破損リスクを低減させつつ、好奇心の赴くままに行動することで、実世界を構造的に理解するロボットが実現できると考えている。

## 7. 参考文献

- [1] Thomas Wisspeintner, Tijn van der Zant, Luca Iocchi and Stefan Schiffer, “RoboCup@Home: Scientific Competition and Benchmarking for Domestic Service Robots,” *Integration Studies*, Volume 10, Issue 3, pp. 392-426, 2009.
- [2] Begus, Katarina, Teodora Gliga, and Victoria Southgate, “Infants learn what they want to learn: Responding to infant pointing leads to superior learning,” *PLoS One* 9(10): e108817, 2014.
- [3] Mitsuhiro Shikida and Koji Asano, “A Flexible Transparent Touch Panel Based on Ionic Liquid Channel,” *IEEE Sensors Journal*, Volume 13, Issue 9, pp.3490-3495, 2012.
- [4] Tomoaki Nakamura, Takayuki Nagai and Naoto Iwahashi, “Grounding of Word Meanings in Multimodal Concepts Using LDA,” in *Proceedings of 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3943-3948. 2009.
- [5] Aaron van den Oord, Oriol Vinyals and Koray Kavukcuoglu, “Neural Discrete Representation Learning,” arXiv:1711.00937, 2017.
- [6] Tuomas Haarnoja, Aurick Zhou, Kristian Hartikainen, George Tucker, Schoon Ha, Jie Tan, Vikash Kumar, Henry Zhu, Abhishek Gupta, Pieter Abbeel and Sergey Levine, “Soft Actor-Critic Algorithms and Applications,” arXiv: 1812.05905, 2019.
- [7] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9, 2015.
- [8] Todorov, Emanuel and Erez, Tom and Tassa, Yuval, “MuJoCo: A physics engine for model-based control,” in *Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.5026-5033, 2012.

## 謝辞

本研究を進めるにあたり、ご指導をいただいたアドバイザー教員の堀井隆斗講師に心よりお礼申し上げます。また、ロボットに配置する触覚センサを選定するにあたり、快く相談に応じていただいた川節拓実助教に深く感謝いたします。