

感覚運動予測学習による物体形状表現の獲得モデル

野口 渉†

北海道大学 大学院情報科学研究院†

1 はじめに

我々は外界の空間の構造を認識した上で、空間に存在する物体の大きさや形状を認識することができる。一方、外界を観測する感覚刺激そのものには、その背後にある空間の構造は明示的に示されず、空間構造に基づく外界の認識は感覚刺激の単純なマッピングでは実現されないと考えられる。

我々の外界についての認知は単に受動的に感覚刺激を受け取るだけではなく、自らの運動に応じて感覚刺激が変化し、感覚運動の依存関係を経験・学習することによって構築されていくという説がある[1]。また、感覚運動の依存関係の学習は、脳の機能を統一的に説明するとして注目される予測処理仮説の枠組みによっても説明できるとされる[2]。Laflaquièreらは、ニューラルネットワークモデルに予測学習という形で感覚運動の依存関係を学習させるシミュレーションを行い、感覚運動経験における予測学習により外界の空間構造の内部表現が獲得されることを示した[3]。一方、彼らのシミュレーションにおいては、外界の空間的構造・空間的位置関係の認識のみが実現され、空間に存在する物体形状等の認識の獲得は実現されていない。

本研究では、感覚運動経験の予測学習を通して、空間構造に加え、物体形状の内部表現を獲得するニューラルネットワークモデルを提案する。

2 シミュレーション・モデル

本研究では、シミュレーション空間中のアームロボットの触覚・運動経験を学習することで、空間構造および物体の形状の内部表現を獲得するニューラルネットワークモデルの構築を試みる。提案するモデルはLaflaquièreらのモデル[3]をベースとしているが、物体が存在する環境において予測を行うモデルに拡張する。

2.1 シミュレーション環境

図1にシミュレーション環境を示す。環境は2

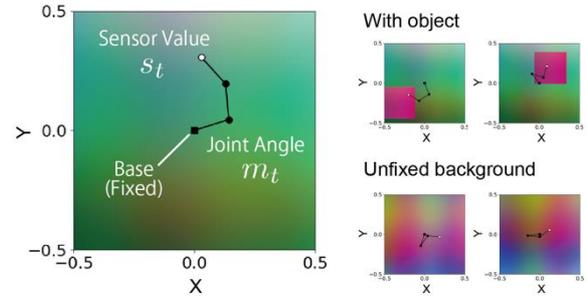


図1 シミュレーション環境とアームロボット

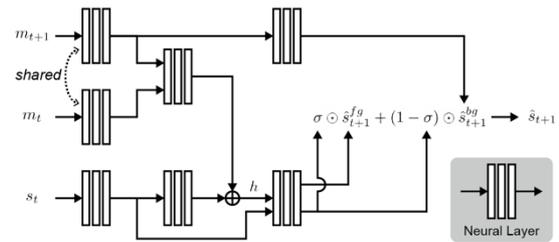


図2 予測ネットワーク

次元平面で、アームロボットは3つの関節自由度をもつ。センサーを備えたアーム先端はアームの関節角度 m_t に応じて2次元平面中を移動する。環境中で固定な背景は触覚を与えるテクスチャを備え、ロボットはアーム先端の座標 p に存在するテクスチャ点の触覚入力 s_t を得る。

ロボットは異なる3つの環境条件を経験する。すなわち、常に位置が固定な背景のみが存在する場合、固定の背景に加え位置が可変な物体が存在する場合、空間を平行移動する非固定な背景が存在する場合、である。物体と非固定な背景も、固定の背景と同様に触覚を与えるテクスチャを備えるが、それぞれのテクスチャに対応する触覚は異なる。とくに、各テクスチャは位置ごとに固有な触覚刺激を与え、1点での触覚から異なる背景・物体を識別可能である。テクスチャは先行研究[3]と同様な方法で生成する。

2.2 予測ネットワーク

上記のアームロボットの触覚・運動感覚において、予測学習を行うニューラルネットワークモデルを構築する。とくに、ロボットの現在の関

Development of object shape representation by sensorimotor prediction learning

† Wataru Noguchi

Faculty of Information Science and Technology,
Hokkaido University

節角度 m_t および触覚 s_t と、次時刻の関節角度 m_{t+1} が与えられたとき、次時刻の触覚 s_{t+1} を予測するネットワークを構築する。アームの関節角度と触覚入力を得る空間座標との対応は多対1で、その対応はモデルにとって未知である。先行研究では、空間中に物体が存在しない条件下で、予測学習によって関節角度と空間座標の対応が構築されていくことが示されている[3]。

図2に予測ネットワークの概略を示す。予測ネットワークは、2つの中間の触覚予測 \hat{s}_{t+1}^{bg} および \hat{s}_{t+1}^{fg} を $(0, 1)$ の範囲の値をとるマスク値 σ によって重みづけたのちに加算することでネットワーク全体の触覚予測 \hat{s}_{t+1} を出力する。中間予測のうち、 \hat{s}_{t+1}^{bg} は次時刻の運動入力のみを用いて生成される。一方、 \hat{s}_{t+1}^{fg} は現在の触覚・運動、次時刻の運動全ての入力を用いて生成される。マスク値 σ もネットワーク内部で生成される。また、現時刻、次時刻の運動入力は共通のエンコーダーで処理され、運動入力と触覚入力は別々にエンコードされたのちに加算される構造になっている。

3 実験

3.1 実験設定

予測学習のために、現時刻、次時刻の触覚・運動入力のペア (m_t, s_t) 、 (m_{t+1}, s_{t+1}) を収集する。アームの関節角度を各時刻で独立にランダムにサンプルし、対応する触覚入力と合わせてペアを作成する。固定背景のみの場合、物体が存在する場合、非固定背景の場合でそれぞれ 100,000 の触覚・運動ペア、合計 300,000 ペアを作成した。また、物体および非固定背景の環境中での変位は、一つの現時刻、次時刻の触覚・運動ペア内では同一であるが、異なるペア間では異なる。また、触覚入力は8次元の値として取得する。

予測ネットワークは、触覚予測の予測誤差の最小化を学習する。最適化アルゴリズムは Adam を用い、50 エポックの学習を行なった。

3.2 実験結果

予測学習後にネットワーク内部の中間ベクトル (図2中の h) を用いて触覚予測を可視化した。図3に可視化した触覚予測を示す。ここでは物体が存在する場合の予測結果についてのみ示している。このとき、ネットワークの中間ベクトルの第1、第2主成分空間上で、中間ベクトルの点を同ベクトルに対応する予測出力を用いて色付けた。色付けは予測出力の8次元のうち3次元をRGB値として用いた。また、ある時刻の触覚・運動ペアに対し、アームの可動範囲を十分密に埋め尽くすように次時刻の運動をサンプル

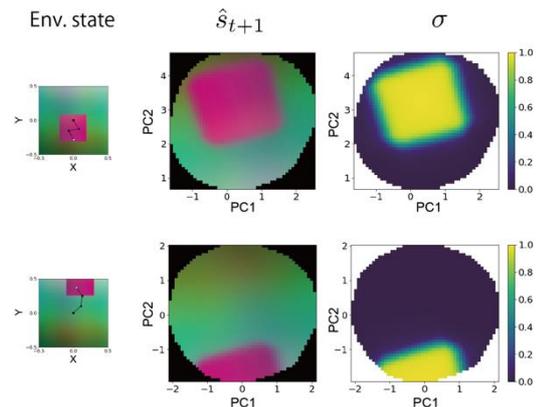


図3 学習後ネットワークの予測の可視化

し組み合わせることで複数の入力 (m_t, s_t, m_{t+1}) を作成し、対応する出力を可視化している。可視化結果からは、主成分空間に実際の空間と同様な物体形状が確認できる。また、ネットワーク内部のマスク出力についても、同じく中間ベクトルの主成分空間に、マスク値のヒートマップとして可視化した。結果として、物体が存在する領域でのマスクの値は1、それ以外の領域では0になっていた。つまり、物体と背景によってマスクを分けるという形で、物体の形状が表現されており、これは、中間ベクトルでの空間構造の表現と同時に獲得されたものである。

4 おわりに

本研究では、空間構造について事前知識をもたないモデルが、予測学習を通して、空間構造および物体の形状の表現を獲得するシミュレーションを行なった。本研究では、物体と背景それぞれが固有なテクスチャをもつことを仮定しており、局所的な情報のみからそれぞれを識別可能であったが、今後は、局所的な情報では物体を識別できない条件においても、物体形状の表現を獲得可能なモデルの構築を試みる。

謝辞

本研究は JSPS 科研費 JP20K19880 の助成を受けたものです。

参考文献

- [1] O'regan, J. K. and Noë, A.: A sensorimotor account of vision and visual consciousness, *Behav. Brain Sci.*, Vol.24, No.5, pp.939-973 (2001).
- [2] Seth, A. K.: A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia, *Cogn. Neurosci.*, Vol.5, No.2, pp.97-118 (2014).
- [3] Laflaquière, A., and Ortiz, M. G.: Unsupervised emergence of egocentric spatial structure from sensorimotor prediction, *Adv. Neural Inf. Process. Syst.* Vol.32, pp.7158-7168 (2019).