F-045

共有モジュールネットワークによる記号操作の身体化
Embodiment of symbolic manipulation using shared module network

嶋田 泰大† 野口 渉‡ 飯塚 博幸‡* 山本 雅人‡*
Yasuhiro Shimada Wataru Noguchi Hiroyuki Iizuka Masahito Yamamoto

1. はじめに

人間は実空間での経験と抽象的な記号操作を相互に対応づけることができる．例えば，2 つのリンゴが足し合わされることを 1+1 という計算の例として理解できる一方，10+3 に対応する状況をイメージすることができる．

ロボティクス等の分野では，身体を持った主体が環境との相互作用を通してボトムアップに概念を獲得する，という身体性が重要視されている[1]．しかし，これらはボトムアップに概念の獲得をすることに限られる．つまり，冒頭の例でいう実空間での経験の抽象化は考えられてきたが，抽象的な操作を具体化して理解することはあまり考えられてこなかった．

一方，筆者らは共有モジュールを用いて実空間での経験と抽象的な記号操作を双方向的に対応づける深層学習モデルを提案している．共有モジュールを用いることで，エージェントがシミュレーション環境を動き回るという実空間での経験と，環境が抽象化された地図上でのナビゲーション操作を統一的に扱えることが示された[2]．この研究において，まずモデルは実空間での経験における視覚の予測学習を行うことで空間の内部表現を獲得する．地図に関する経験においても同様な予測学習により内部表現を獲得するが，共有モジュールを用いることで実空間と抽象的な地図の経験の内部状態が自然に対応づけられる．それによって，地図の情報を頼りに実空間での移動を行えるようになるということを示した．つまり，従来考えられてきたボトムアップな概念の獲得に加え，抽象化された地図からのトップダウンなナビゲーションを実現した．ただし，[2]では離散的な記号操作は扱われていない．

本研究では[2]を拡張し，数に関する実空間での経験と計算という抽象的な記号操作が共有モジュールの利用によって統一的に扱えることを示す．この中で，記号操作が実空間での動作と同じように扱われる，つまり身体化されることを示す．また，記号操作の身体化により，新たな記号操作を学習した時にも対応する実空間での動作として理解できるようになることを示す．

2. シミュレーション

上記を実現するためのシミュレーションについて説明する．まず，シミュレーション環境内にエージェントを用意する．シミュレーションした実空間における経験の中で，

†北海道大学 大学院情報科学院
Graduate School of Information Science and Technology, Hokkaido University
‡北海道大学 大学院情報科学研究院
Faculty of Information Science and Technology, Hokkaido University
*北海道大学 人間知・脳 AI 研究センター
Center for Human Nature, Artificial Intelligence, and Neuroscience, Hokkaido University

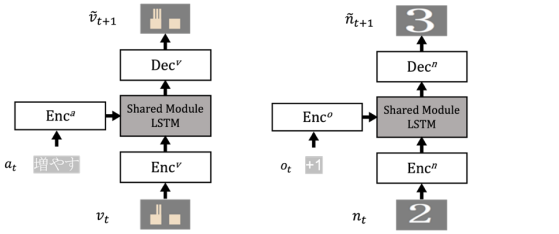

図 1 左：実空間でのモデル．右：抽象空間でのモデル．

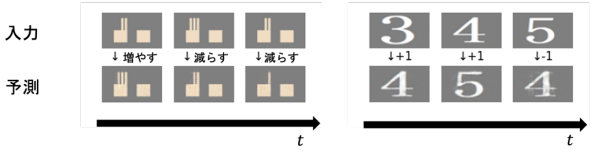

図 2 予測結果．左：「1 つ増やす・減らす」．
右：「+1・−1 」．

エージェントは視覚情報と動作情報を取得する．また，抽象的な記号操作の経験の中で，視覚情報と記号操作の情報を取得する．これらの経験を用いて，3 節で述べるニューラルネットワークで構成された提案モデルにより予測学習を行う．本研究では，人が手や物を用いた実空間での経験によって数の順序の概念を獲得した後に，それをもとに計算という抽象的な記号操作を学習する，という仮定のもと以下の実験を行う．

2.1 実空間でのシミュレーション

数の順序の概念を獲得するための実験として，指を 1 本ずつ曲げ伸ばしした結果，手の状態がどのように変わるかを予測する様子をシミュレーションする．エージェントは現在の手の状態と動作をもとに，次の手の状態を予測する．同様に星，赤玉，白玉を 1 つずつ置く・取る(増やす・減らす)という動作をした場合のシミュレーションも行い，計 4 種類学習を行う．なお，本研究では動作の結果のみを観測するものとしており，動作の途中の視覚は考慮していない．

2.2 抽象空間でのシミュレーション

次に，記号操作においても予測学習を行う．数字は画像として与えられ，与えられた記号操作によって別の数字に切り替わる．エージェントは，現在見えている数字に対して「+1・−1」という記号操作を行った結果，どのような数字になるかを画像として予測する．

3. モデル

実空間での経験と抽象空間での操作を対応づけるために，共有モジュールを用いたニューラルネットワークモデルを構築する．図 1 にネットワーク構造を示す．図 1 左は実空間での学習を行うネットワークである．入力として現在の視覚情報 v_t と，次に行う動作 a_t が与えられる．それぞれがエンコーダによって処理された後に LSTM に入力され，デコーダを通して予測した視覚情報 $\tilde{v}_t$ を出力する．なお，

エンコーダとデコーダはどちらもフィードフォワードであり，全結合層と CNN で構成されている．

実空間での学習が終わった後に図 1 右のモデルを用いて抽象空間での学習を行う．入力としては現在の数字 n_t とそれに対して適用する演算 o_t が与えられ，次の数字 $\tilde{n}_t$ を画像として予測する．この時，実空間での経験を学習した LSTM を共有して使用する．これが共有モジュールである．なお，抽象空間での学習の際には LSTM 以外のモジュールは別な新しいネットワークを用いる．

学習データとして，実空間での経験の系列と抽象的操作について 1 系列 40 ステップとして，それぞれ 300 系列ずつ用意する．また，視覚情報は画像として与えられ，実世界での動作，抽象的操作はそれぞれランダムな 5 次元のベクトルとして与えられる．なお，これらにはガウシアンノイズが付加され，それぞれエンコーダ Enc^a および Enc^o により 2 次元ベクトルに変換される．また，視覚エンコーダの出力をある確率で全てゼロベクトルにするというマスク処理を行う．これは，人が目を瞑りながら動作の情報だけで手や物の状態の予測を行う様子に対応する．これは，記号操作に関しても同様である．また，損失関数としてはデコーダの出力画像と正解の画像との 2 乗誤差を用いる．

4. 共有モジュールを用いた具体・抽象の対応づけ

実空間，抽象空間それぞれでの次状態予測の結果を図 2 に示す．それぞれ動作の結果，抽象的操作の結果を正しく予測できていることがわかる． ここで，共有モジュールの内部状態を PCA で 2 次元に可視化したものが図 3 左である．薄色部は実空間での経験の際に遷移していく内部状態を表している．色分けは，出力に対応する数字によって行われている．図に書かれている数字の通り，出力が 0 の時から 10 の時まで内部状態が横軸方向に連続的に分布している．ここで，薄色部は 4 つの系列に分かれており，図に示す通り 4 種類のオブジェクトに対応している．ここで，例えば 1 番上の系列は手を用いた経験をしている時の内部状態である．黒矢印は，「1 つ増やす」動作を行なったことによって内部状態が遷移する様子を示す．以上に関して，4 種類の系列が同じ方向に並び，同じ色が同一直線上に乗るような形で内部状態が遷移していることから，数の順序に関する概念が形成されたと考えられる．次に，濃色部は抽象空間での学習の際に遷移していく内部状態である．これを見ると濃色部の並びが薄色部に対応していることがわかる．これは，「+1・−1」という記号操作による内部状態の遷移と「1 つ増やす・減らす」動作による内部状態の遷移がそれぞれ対応するように学習されたことを表している．

ここで，図 3 左に実空間での動作を Enc^a に入力した時の出力(青)と抽象的な記号操作を Enc^o に入力した時の出力(オレンジ)を示す．左上の分布が「1 つ増やす」と「+1」 をそれぞれエンコードした時の出力に対応し，右下の分布が「1 つ減らす」と「−1」に対応する．この結果を見ると，エンコードされた実動作に対して，エンコードされた抽象的な操作が後から近づくように学習されたことがわかる．つまり，実動作と同じように解釈され，記号操作が身体化されたといえる．このような身体化によって，上記のような内部状態の対応が実現されたと考えられる．

5. 新たな抽象操作の学習

ここで，新たな抽象操作「+2・−2」を学習する．モデルは「+1・−1」を学習したものと同じだが，Enc^o 以外は学習

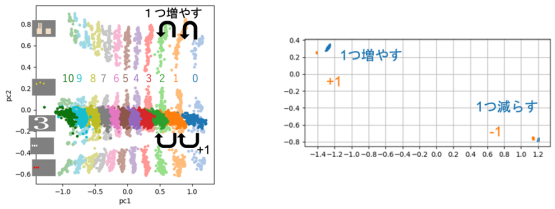

図 3　左：共有モジュールの内部状態．
右：Enc^a と Enc^o の出力値．

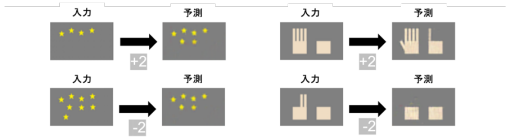

図 4　「2 つ増やす・減らす」の予測結果．

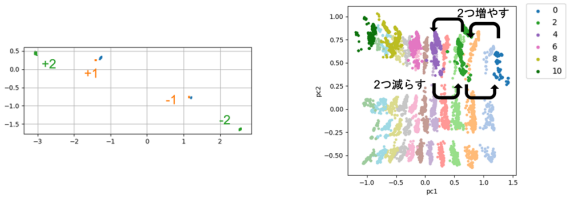

図 5　左：Enc^a と Enc^o の出力値(+2・ −2 の場合)．
右：共有モジュールの内部状態(2 つ増やす・減らす)．

せずパラメータを固定する．学習後の Enc^o の出力を図 5 左に示す．緑が「+2・−2」に対応する．青とオレンジは図 3 左と同じである．これを見ると，「+2・−2」は「+1・−1」とは 異なった点として学習されていることがわかる．

次に，図 1 左の Enc^a を新たに学習した Enc^o に入れ替えたモデルを用いて，実空間での予測ができるかを確認する．つまり，「+2・−2」を入力として得た時の Enc^o の出力を用いて実空間での予測をする．その結果が図 4 である．2 つ増えた状態と 2 つ減った状態を予測できていることがわかる．その時の内部状態の遷移が図 5 右である．ここから，図 3 では内部状態が 1 つずつ遷移していたのに対し，1 つ飛ばして遷移している様子がわかる．以上から，抽象的な記号操作「+2・−2」が身体化され，モデルが「2 つ増やす・減らす」という新たな動作として解釈できるようになっていることが確認できた．

6. おわりに

本論文では，抽象的な記号操作の身体化を共有モジュールを用いることで実現した．また，それをエンコーダの出力の対応，内部状態の対応という形で確認した．以上により，新たな記号操作も身体化され，対応する実空間での動作として解釈されるようになることを確認した．ただし，今回の実験設定は簡略化されたものである．今後，より現実に近い設定にしていくことが必要である．

謝辞

本研究は JSPS 科研費 JP22H03909 の助成を受けたものです．

参考文献

[1] Pfeifer, Rolf, and Christian Scheier. “*Understanding intelligence*”, MIT press (2001)
[2] Noguchi, Wataru, Hiroyuki Iizuka, and Masahito Yamamoto, "Multi-modal shared module that enables the bottom-up formation of map representation and top-down map reading", *Advanced Robotics* 36.1-2 (2022)