

[研究助成 (A)]

海馬系の脳計測との融合を目指した人工知能技術の開発

2251032

研究代表者
(助成金受領者)株式会社国際電気通信基礎技術
研究所

連携研究員

細 谷 晴 夫

共同研究者

国際電気通信基礎技術研究所

室長

Aurelio Cortese

[研究の目的]

AI-脳融合とは、脳における情報処理の一部をAIに肩代わりさせることにより、脳からの情報（fMRIなどで計測した脳活動）を高精度に解釈して可視化したりする技術である。本研究では長期目標として、脳の海馬系の情報処理方式に倣ったAIモデルを利用することにより、脳から自分のいる場所などに関する脳情報を高精度に読み取る技術の確立を目指している（図1）。海馬系は、過去に体験した事象のスナップショットのような記憶（エピソード記憶）を司る脳の部位であり、場所や動作に関する記憶、すなわち「認知マップ」はその典型例である。近年では物理空間だけでなく概念空間でも認知マップが形成されることがわかっている。上記のような長期目標においては、そのような海馬

系の情報処理方式を模倣したAIモデルを構築することが、中核的な技術課題となる。しかし、海馬系に倣ったAI研究は未開拓で、強い制約条件下でのみ動作する神経回路モデルがあるに過ぎない。したがって本研究では、AI-脳融合応用での使用にたえるスケーラブルな海馬系AIモデルを開発することを目的とする。

[研究の内容、成果]

本研究の主目標は、AI-脳融合応用を見据えた海馬系のAIモデルの構築を行うことである。具体的には、①海馬系の神経反応特性を高い精度で再現できる、②学習効率が高くスケールアップする、③学習方式が高い汎化性を持つものを目指す。特に、モデルでの訓練にはない未知の空間への汎化性は応用上で重要である。そのようなモデルの考案にあたり、大きな方針として、海馬系が回路構造の異なる海馬と嗅内皮質という2つの領野の連関計算をしているという神経科学の知見からヒントを得て、マップの幾何構造を表現する「抽象表現」と、個別のマップを表現する「具象表現」に分離するという考え方をとることにする。これにより学習方式の汎化性を担保するのである。

まず、嗅内皮質のグリッド細胞（Hafting et al. 2005）を模した抽象表現を仮定し、それを具象化する形でタスク背後の認知マップを構成するモデルを考案した（図2A）。具体的には、

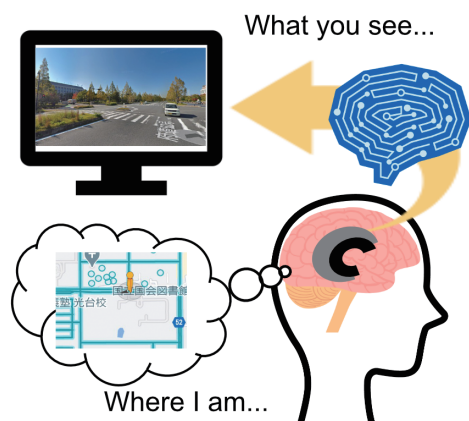


図1 海馬系の脳 AI 融合技術

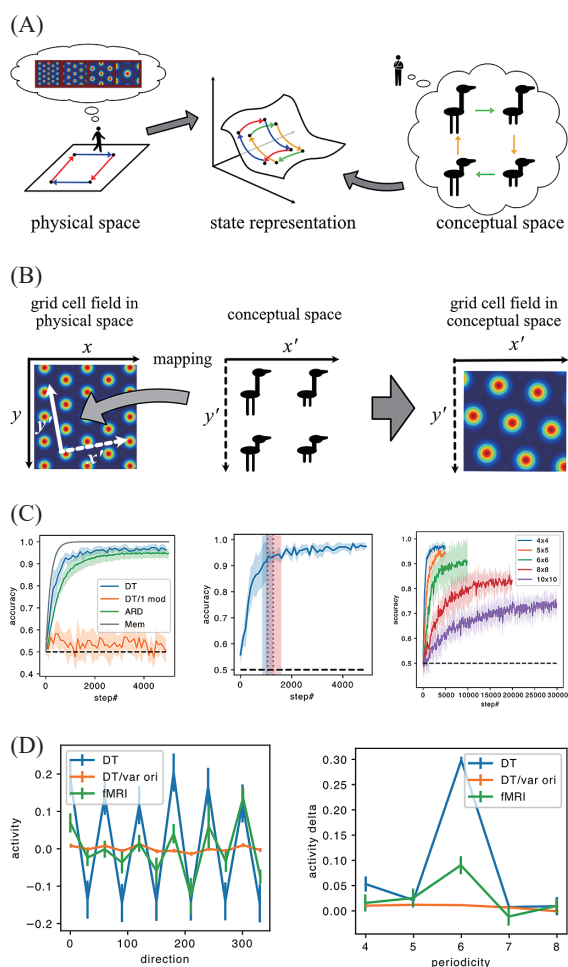


図2 概念空間マップを学習する海馬系 AI モデル

グリッド細胞集団の活動が形成する低次元マニフォールドに着目し、その幾何学的構造をダイナミクスとして表現して再利用することで、グリッド細胞が表す物理空間から概念空間へのドメイン間転移を実現する (図 2B)。タスクとしては、ある対象 (画像など) の集合が 2 次元概念空間に配置されていると仮定し、対象ペアごとの関係を推論する試行を繰り返すものを用いた。それにより、対象集合の配置とマップ構造を推定することが目的である。タスクは離散的 (格子状の) 空間もしくは連続的空間を仮定する二種類を用いた。

モデルの学習では、各対象をマニフォールド上の状態パラメタとして、また対象間の関係を状態間の変位パラメタとする。そして、試行の平均成功確率を目的関数として最適化する。そ

の際、すべての状態パラメタが常にマニフォールド上にとどまるよう制約付き最適化を行う。最適化手法として、「manifold trekking」と呼ぶ独自技術を開発した。

この手法では遷移関数を用いて状態を更新することで、幾何構造を保ったまま学習を進め、概念空間の構造がグリッド細胞のユークリッド幾何に整合する形で自動的に埋め込まれる。

モデルを実行した結果、いずれの課題でも数千ステップ以内に高精度に到達した (図 2C 左)。また、未学習の組み合わせに対しても推移的推論が可能で、これは過去の実験で示されている人間被験者の速度と同程度だった (図 2C 中)。これは、単なる記憶ではなく、潜在的な幾何構造を学習していることも示している。しかもある程度タスクサイズを大きくしても学習可能だった (図 2C 右)。さらに、概念空間内の移動方向に対する平均活動を解析したところ、60 度周期の明瞭な六角変調が再現された。この結果はヒト fMRI 研究 (Park et al. 2021; Constantinescu et al. 2016) で報告された現象と整合している (図 2D)。これは概念空間でのグリッド細胞の出現を計算論的に説明するものである。特に、概念空間と物理空間の間にはほぼ直交的な線形写像が形成されたとき、最も安定した六角変調が得られた。また、グリッド細胞のモジュール (周波数の種類が異なるグリッド細胞の集団) の個数を減らすと性能が低下し、単一モジュールでは学習が破綻したことから、複数周波数構造が精密な位置符号化と転移に不可欠であることが示唆された。モジュール間のグリッド方位が揃っていない場合には六角変調が消失し、これは動物実験 (Stensola et al. 2012) で観察されるモジュール方位整合性と一致する。さらに、本枠組みを三次元概念空間へ拡張する試みとして、二つの独立したグリッド細胞ネットワークを組み合わせる手法を導入した。その結果、より高次元の構造に対しても安定した学習が可能であることが示された。これは、複数の二次元表現の組み合わせによって高

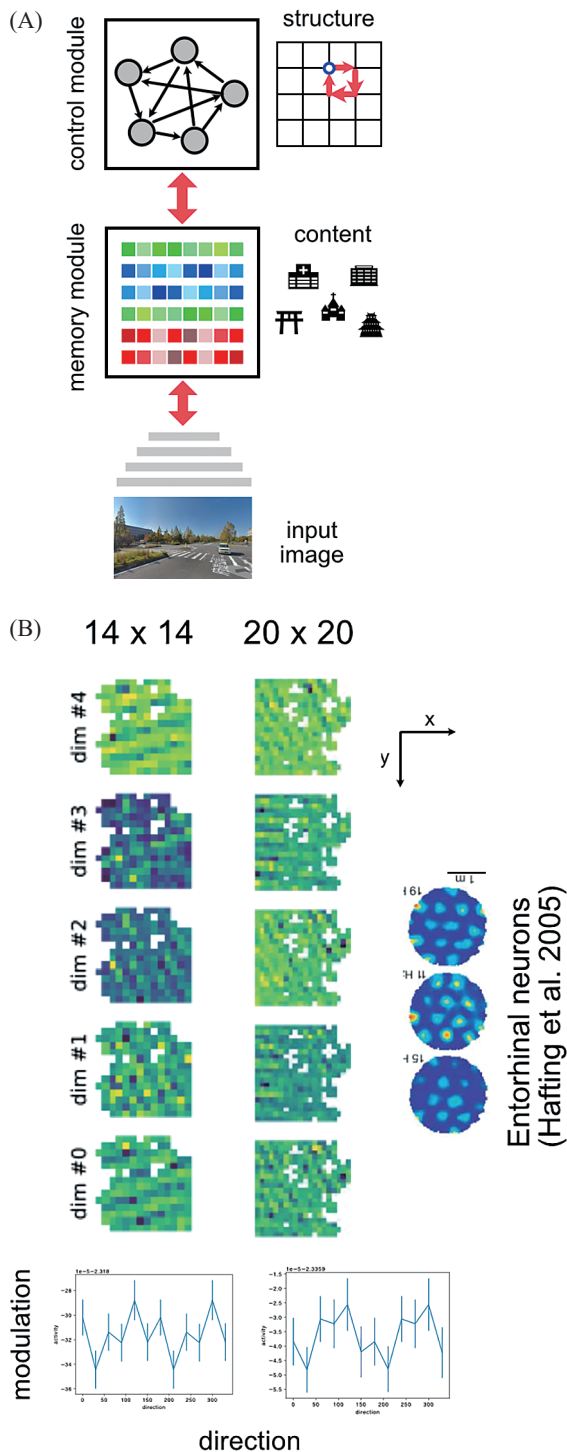


図3 移動探索型の海馬系 AI モデル

次元空間を表現できる可能性を示唆する。

本研究では上記の成果に加え、物理空間を移動探索しながら認知マップの一般的な構造を学習する AI モデルの構築も試みた (図 3A)。このモデルでは、具象表現 (海馬) に加え、抽象

表現 (嗅内皮質) も学習するような「抽象・具象分離アーキテクチャ」を導入した。このアーキテクチャは、本研究代表者が過去に開発した意思決定モデル (Hosoya, ICLR 2024) を出発点とし、読み書き可能なメモリモジュールと、リカレントネットによる制御モジュールを組み合わせ、全体を end-to-end で学習する。

このようなアーキテクチャにより、具体的タスク環境に依存しない「認知マップの一般構造」を学習でき、これにより新規のシーンを含む未知のタスク環境への汎化性が担保できると考えられる。近年類似する神経回路モデルによって、哺乳類の海馬系の神経反応特性が再現できるという結果が発表された (Whittington+, 2020) が、詳細な作り込みが多数含まれ、学習効率は非常に悪く、小規模で人工的なトイタスクでしか動作しない。特に、入力の記憶創生・想起を行うメモリモジュールの学習効率にボトルネックがあり、メモリサイズの 4 乗の計算量がかかっていた。一方、本研究では、2 乗の計算量で抑えたメカニズムを用いることで解決を図っている。

図 3B に構築したモデルの実行例を示す。タスク環境のサイズ 14×14 及び 20×20 について実行した。図 3B の上側に、モデルのユニット 5 個について可視化した。それらを見ると、規則的な構造を持つパターンが見られ、それらが哺乳類の海馬系を構成する嗅内皮質の神経細胞の持つ六角格子パターン (図 3B 右側) と類似していた。定量化するために、図 3B 下側に図 2 と同様な方法で方位に関する変調をプロットしてみると、60 度ごとの周期的パターンがあり、特に 14×14 の場合には明確にある。しかしながらまだ神経応答パターンの再現性が十分とはいえず、かつこれらのモデル訓練には 1 日以上のかかることから、今後更なるモデルの改良が課題となる。

以上より、本研究では抽象・具象表現の分離という考え方から、海馬系における概念空間マップの獲得のモデルと、抽象表現 (グリッド

細胞) 自体を学習するモデルの開発を行った。前者のモデルに関しては、本研究の目標であった神経特性の再現性、効率性、汎化性の3つに関してある程度達成しているものの、後者のモデルに関しては目標が予想よりも遥かに難題であることがわかり、将来課題が多く残る格好になった。しかしながら、本研究を通じてニューラルマニフォールド上での制約付き最適化という枠組みが、神経活動の幾何構造と下流計算と

の関係を明示的に扱う一般性がある可能性があり、将来性を醸し出す研究内容になったとも言える。

[成果の発表, 論文など]

細谷 晴夫. グリッド細胞の再利用によって抽象関係表現を学習する構造類推モデル. 第48回日本神経科学大会, 2025. 口頭発表.

Haruo Hosoya. Cross-domain transfer of spatial representations through grid-cell geometry. In preparation.