

## AGI 開発をナビゲートするための能力マップ

## A Capability Map for Guiding AGI Developments

大澤正彦 \*1\*2

Masahiko Osawa

大森隆司 \*3\*6

Takashi Omori

高橋恒一 \*4\*6

Koichi Takahashi

荒川直哉 \*5\*6

Naoya Arakawa

坂井尚行 \*5\*6

Naoyuki Sakai

上野道彦 \*5\*6

Michihiko Ueno

今井倫太 \*1

Michita Imai

山川宏 \*5\*6

Hiroshi Yamakawa

\*1 慶應義塾大学大学院理工学研究科

Keio University, Graduate School of Science and Technology

\*2 日本学術振興会特別研究員 (DC1)

Japan Society for the Promotion Science, Research Fellow (DC1)

\*3 玉川大学 工学部

Tamagawa University, College of Engineering

\*4 理化学研究所 生命機能科学研究センター

RIKEN Center for Biosystems Dynamics Research

\*5 ドワンゴ人工知能研究所

Dwango Artificial Intelligence Laboratory

\*6 全脳アーキテクチャ・イニシアティブ

The Whole Brain Architecture Initiative

本研究では、人レベル AGI の開発を効率的に行うためのガイドとしての能力マップを新たに提案する。能力マップは、人レベル AGI に必要とされる能力を階層的に整理し研究の現状を客観的に評価することで、重点的な研究開発が必要な能力を明確化することができる。さらに、能力マップにおいて階層化した個々の能力に対して、それを実現する内部実装、評価する工学的タスク、そして対応する脳領域をそれぞれ割り当てる。特に能力に脳領域を結びつけることは、同じ脳領域に対応づけられた複数の能力を統一的なメカニズムで実現する可能性を示唆し、AGI において重要な汎用性の向上を促進させると期待する。能力マップを AGI 研究開発コミュニティの開発サイクルの中心として位置付けることで、研究活性化が可能と考えられる。

## 1. はじめに

深層学習の成功をきっかけに、人間レベルの汎用人工知能 (AGI) の実現への期待が高まっている。人間レベル AGI が実現すれば、多くの産業分野で発展を加速できると期待される。一方で、人間レベル AGI は汎用的であるがゆえに、それを実現するために工学的に実現する必要がある能力 (認知機能) は多岐にわたり、その全体像を把握することは容易ではない。ここで、能力とは、工学的に実現することが期待されるあらゆる粒度の実行機能とする。

したがって今後は、開発済みの AI 技術でカバーできる部分とできない部分を明確化し、体系的にまとめながら技術開発を効率的に進めていく必要がある。

類似のものとして、DARPA が提供している認知機能のリストがある [DARPA 05]。しかし、リスト化だけでは直近の課題の明確化はむずかしい。実装する必要がある能力 (認知機能) を要素に分解・階層化して、研究の現状と結び付けて提供することが重要であろう。

そこで本研究では、人間レベル AGI 実現に向けて実装すべき能力を分解・階層化した能力マップの構築を提案する。能力マップを作成する主な目的は以下の 4 つである。

- 人間レベルの AGI の全体像を踏まえた、開発に必要な構成要素の洗い出し
- 直近で着手可能な、あるいは着手すべき技術課題の可視化
- 現在の研究進捗状況や、過去からの技術発展からの差分を定量化

- 技術的な現状や見通しを踏まえた、ロードマップの材料として利用

能力マップはノードとエッジからなるグラフとして表現される。1 つのノードは、能力、内部実装、タスクリスト、対応する脳領域がセットになっており、エッジは能力の階層関係を表現する。能力と内部実装が対応づくことで、工学的な手法と AGI に組み込まれた場合の位置付けを示すことができる。マップを構成する個々の能力は、それを含むエージェントが解くべき工学的タスクと対応づけることで、能力の実現度を定量化することができる。そして、各能力に主に寄与する脳領域や神経回路を対応づけることで、同一のメカニズムでクリア可能な、あるいは関係があると考えられるタスク群の可能性を絞ることや、モジュールの再利用可能性を示唆できる。

## 2. 能力マップ

## 2.1 能力マップの概要

提案する能力マップとは、人間レベル AGI の構築を戦略的に行うために、AGI 開発コミュニティで知識共有するためのツールである。図 2 にナビゲーション能力を最上位ノードにおいた場合の能力マップの一例を示す。能力マップは知的な能力をノードとし、それらの依存関係をエッジとした有向グラフで表現される。

能力マップは能力の他に 3 つの要素を 1 組にしてノードとする。

**内部実装** 対応する能力を実現するための工学的なメカニズム。

**タスクリスト** 達成することが求められる工学的タスクのリスト。タスクリストを能力と対応づけることで、その能力がどの程度達成されたかを定量的に評価することができる。さらにある能力の要素への分解をタスクから裏打ちすることで、能力の階層性の妥当性を向上させる。

連絡先: 大澤正彦, 慶大/学振 (DC1), 〒 223-8522 神奈川県横浜市港北区日吉 3-14-1, 045-566-1749, mosawa@ailab.ics.keio.ac.jp

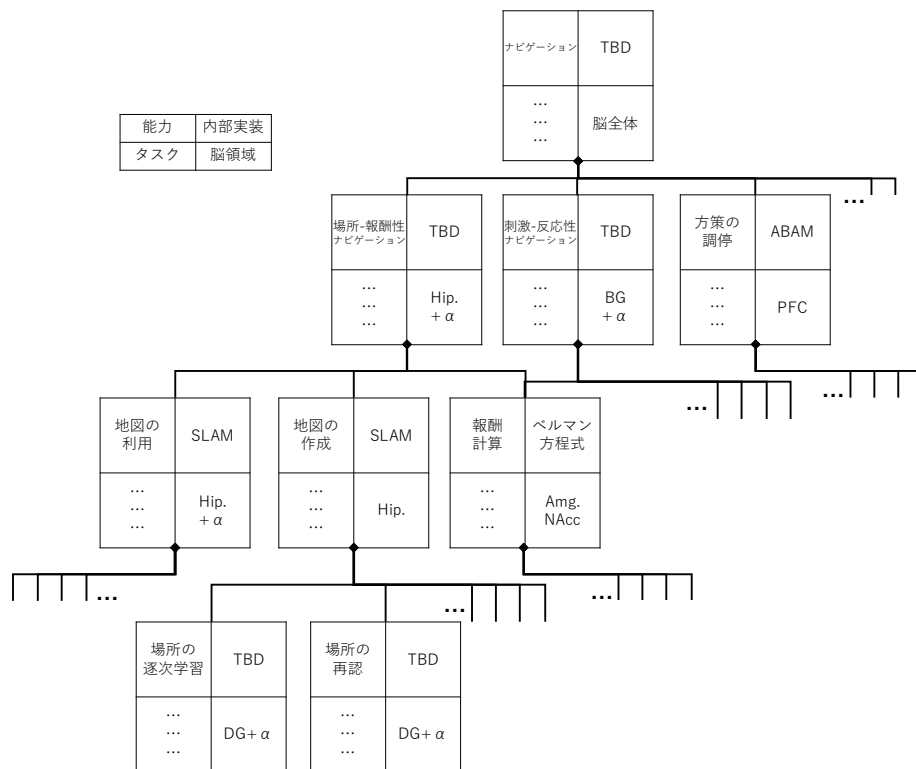


図 1: 能力マップの例. 最上位のノードにナビゲーションをおいて構築した例. 1つのノードは複数の上位もしくは下位ノードを持つ. 内部実装, タスクリスト, 脳領域はいずれも完全に埋めることは困難で, これら3つを埋めることで AGI 研究が進展すると考えられる.

対応する脳領域 能力を実現するために中心的な役割を担っている脳領域や特定の神経回路を対応づけることで, 最終的に目指す人工知能の汎用性を向上させる. 詳細は 2.3 で説明する.

## 2.2 能力マップの利用例

先に述べたように, 能力マップの利用は人間レベルの AGI を長期的に研究開発していく一助となると考えている. 本節では現在想定している具体的な利用方法について列挙する.

### タスク設定とタスク解決のサイクル

これまでの AI 研究では, タスクを設定して定式化することは新たな機能の実現に寄与して来た. 能力マップにおいては 1つの能力に対して 1つ以上のタスクを割り当てるが, 適切なタスク設定も AGI 開発コミュニティの中で提案・精緻化されることを想定している. 能力マップ上でタスク設定とタスク解決が同時に取り組まれながら, 能力マップ自体も精緻化されていくというサイクルを実現することで, 人工知能の未解決問題に取り組んでいけると考えられる.

### 神経科学的な知見の体系的整理

各能力は, それ自体またはその下位階層の全ての能力に神経科学的な対応づけがされることを目指す. それにより, 神経科学的知見と認知機能の対応づけに寄与できる可能性がある. さらに, 認知機能として広く認められながらも, その神経基盤が明らかでない機能を明確化することにも貢献しよう.

### 知能研究に体系性を与える参考資料として

人間レベル AGI を開発する目的で開発した能力マップが多くの関係者の合意を取れるなら, 知能研究に体系性を与える参考資料となろう. 特に, タスクリストを用いて能力の体系性の裏付けをするという優位性がある.

## 研究の活性化

能力マップを利用することで, 研究があまり行われていない領域を明確化することができ, 能力マップを分析すると, 研究が進んでいない, あるいは少ない知能の領域を明確化できよう. その場合には, 活性化が必要と考えられる領域に対して研究プロジェクトを発足させることで, 活性化させられる可能性がある. また, 能力を詳細に分割してタスクを定義することで, 解決すべき要件が明確化できるため, 多くの人材が AGI 開発コミュニティに参画しやすくなると期待する.

### ロードマップの作成

能力マップに工数推定とリソース計画を組み込むなら, ロードマップとしても利用することができる. それは, 技術的な課題を体系的にまとめたロードマップとなることから, 抽象的な予測から作るロードマップよりも高い信頼性があるであろう.

## 2.3 脳型制約による汎用性の向上

能力マップでは全ての能力が少なくとも間接的に脳器官に接地していることを目標とする. 能力を脳領域と対応づけることは, 能力マップが表現する知的システムの汎用性を 2つの方法で向上させるきっかけを作ることができる.

### 能力の一本化

別の能力のラベルがはられたノードが, 同一の脳領域を割り当てられていた場合に, 実は同じ能力について記述している可能性を示唆する. これによりグラフのノードが一本化されるため, グラフが小さくなる. 能力マップ自体が洗練されることで, 汎用性に寄与する.

### 内部実装の一本化

別の内部実装によって実現されていた能力ノードが, 同一の脳領域を割り当てられていた場合に, 実は同じ内部実装で実

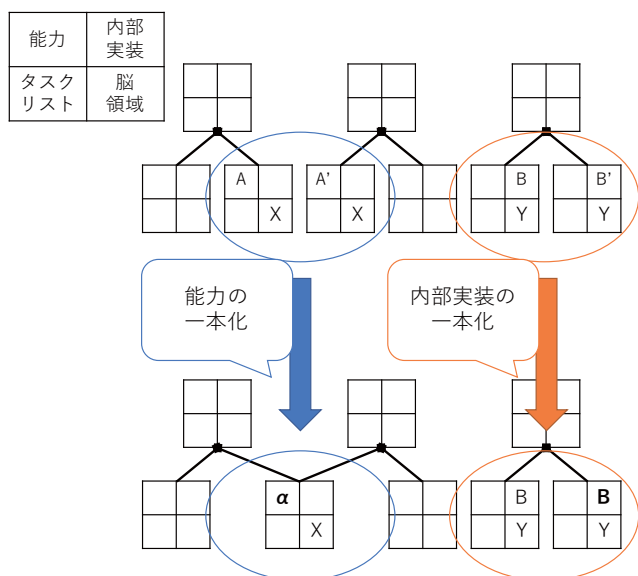


図 2: 脳型制約による汎用性の向上

現可能もしくは実現すべきである可能性を示唆する。能力マップは能力の階層構造によってグラフが形成されるため、内部実装の変更はグラフ構造を直接変化させない。一方でグラフ全体を実現するために利用される技術の数が削減されるため、汎用性・再利用性に寄与する。

#### 2.4 全脳アーキテクチャプロジェクトにおける能力マップの位置付け

全脳アーキテクチャプロジェクト [Yamakawa 16] では、脳全体を AI 技術モジュールの組み合わせとして捉え、モジュールを組み合わせることで人間レベル AGI の実現を目指している。またプロジェクトで参考にする全脳のモデルは最初から人間を想定するよりも、人間と脳のアーキテクチャの共通性が高くありながらも、構造がよりシンプルで神経科学分野における知見の多い動物をマイルストーンとするべきだと考える。具体的には、参考にする脳構造をげっ歯類から、類人猿、そして人間へと推移していく予定である。その過程で動物間に共通する、あるいは異なる認知機能を明確化させることで、げっ歯類から類人猿、類人猿から人間へとプロジェクトの対象をシフトしていく際に、円滑な技術の再利用ができるようにする必要がある。

本研究で提案している能力マップは、段階的に扱う能力を拡張していく全脳アーキテクチャプロジェクトと相性のいい設計になっている。能力マップは目標とする全体像を還元論的に分解して記述する。げっ歯類レベルの認知機能として体系化された能力マップは、類人猿や人間に拡張した際にほとんど同一の階層構造のまま転用できる可能性が高い。

### 3. おわりに

本研究では、AGI 開発をナビゲートするための能力マップを提案した。能力マップは人間レベル AGI に必要な能力が階層的に記述されている。特に汎用性を向上させるために脳との対応づけを行うことで、一本化すべきノードの発見や、共通の内部実装で解決すべきタスク群の候補を選出することができると考えられる。

AGI 開発は今後、長期的なスパンでの研究開発が必要な領域であり、効率化を図るための方法論は今後さらに需要が大き

くなるであろう。今後のアクションプランとしては、提案した能力マップをさらに精緻化して、実際に AGI 研究開発コミュニティにおけるサイクルのかなめとして運用して行くことを目指す。特に能力マップは多くの立場から合意を得られることが重要であるため、ワークショップ形式で議論を進めることを次なるアクションプランとして想定している。

#### 参考文献

- [DARPA 05] DARPA, "BICA, Biologically-inspired cognitive architectures, Proposer Information Pamphlet (PIP) for Broad Agency Announcement 05-18.," Arlington, VA: Defense Advanced Research Projects Agency, Information Processing Technology Office, (2005)
- [Yamakawa 16] Hiroshi Yamakawa, Masahiko Osawa and Yutaka Matsuo: Whole Brain Architecture Approach Is a Feasible Way Toward an Artificial General Intelligence, The 23rd International Conference on Neural Information Processing, pp. 275-281, Kyoto, Japan, (2016).