

応用脳科学研究センター

大森隆司 研究室



研究室が目指すこと

研究室の名称は「認知アーキテクチャ研究室」である。認知アーキテクチャとは、人の脳が実現する認知過程を計算としてとらえ、その処理システムの基本デザイン、すなわちアーキテクチャを解明するという意味で付けた名称である。その名の通り、本研究室は人間の心の振る舞いを脳のメカニズムによる情報処理の過程と想定し、それがどのようなものか、その実現のための神経回路の仕掛けはどういうものか、さらにその裏にある計算過程や理論はどのようなものか、ということを計算論の立場から解明することを目指している。

もちろん、このような大それた試みが簡単に成功する訳もなく、試行錯誤と失敗の連続ではあるが、それなりに多少の進展はある。それを飽きずに続けることが、新たな視点や方法に行き着くのだと信じて続ける。これが当研究室の実態である。

現在、人の心の根幹にあると考え、何らかのモデル化をして研究の俎上に載せたいと思っているのが「価値」である。今日、人工知能（AI）の研究は機械学習の進歩により多くの人々の注目を集めている。その起源は脳の学習システムにあり、脳の知能のモデル化が新しい人工知能を生み出すとも考えられている。しかし、そのモデルを詳しく見ていくと、引っ掛かりを感じた。それは報酬（ゴールともいう）である。古典的なツリー探索から最新の強化学習まで AI の基本機能は探索であり、それを導くのはユーザーが与える報酬（ゴール、誤差ともいう）である。それにより、ユーザーの目的に沿った解を導くのであるが、問題は報酬（価値）を人が任意に決めることにある。単機能の人の道具としての AI ならそれでよいかもしれない。しかし現実の社会は複雑であり、人が適当に定めた報酬に添うはずの行為が想定外の結果を引き起こすこともある。そのとき、単純な価値のみに

駆動される AI には対応するすべがない。原因は、世界の複雑さに見合わず単純な価値設定にあるように見える。

一方で我々は、日常行動では自身の目的だけでなく、複数の立場からその行為の影響を予想し、その中でもっとも効果的かつ副作用の少ない行動を選んでいる。私たちが未来の AI に期待するのは、このような人にも容易ではない意思決定をサポートしてくれる、あるいはそのように行動してくれる AI であろう。その実現のカギとなる要素は「価値」「社会」など現在の AI ではいまだ困難で人に任されている部分のモデル化、ということになる。

それに対して、人をはじめとする多くの動物は、多様な価値の要求に対応する汎用的な知能を実現している。そこに実現されている価値は、生存、繁殖という生物にとって根源的なものから、安心安全、愛着愛情、さらには社会的な価値に関わる金銭的価値、自己実現など極めて幅広い。それらは脳というデバイスの中に組み込まれ、ときに協調し、多くの場合は競合しながら知能の振る舞いを制御し、結果として人の心のような複雑な存在を実現している。そのモデル化なしに人と対等に向き合える AI の設計はないであろう。このような視点から、本研究室では以下の二つのプロジェクトが進行している。

研究内容

1. 価値に駆動された意思決定過程の統合モデル

人の意思決定に関わる理論として、反射（先天的な行動傾向）、強化学習（経験による価値最大行動の強化）、推論（その場での予測と価値評価）の三種を取り上げ、それらを価値探索の状況に応じて切り替える意思決定モデルの構築を志向している（図1）。

この過程で、ヒトで特に強力でとされる論理的／離散的

な推論については、脳のシステム動作としてそれが発生する可能性を検討している。脳は基本的には大量の神経細胞による連続時間連続量のデバイスと考えられるにも関わらず、論理のような離散的な思考を生み出すにはそれなりのしかけが必要である。さらに論理的な推論には意識が何らかの形で関わることはほぼ自明であり、それが説明できるモデルが必要とされる。

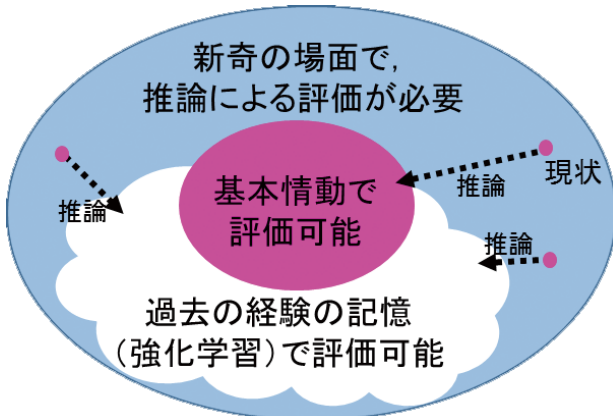


図1 脳の意思決定は複数の価値の統合により幅広い場面に対応している。それらの統合には価値の比較の機構が必要とされる。

研究内容

2. 子どもの行動観察から見る社会的認知能力の発達

社会的認知とは、他者との共同活動を可能とする認知能力であり、他者が考えていることの推定や、他者が考える自分の内部状態の推定など、複雑な推論を含むものと考えられる。このような能力は発達心理学では「心の理論」と呼ばれ、ヒトでは3歳から5歳の就学期の前までに発達するとされる。しかしそのメカニズムの理解はいまだ不十分であり、我々はヒトに匹敵する社会的認知の能力をAIとして実現できる見通しはない。

社会的認知と通常の認知の違いは何なのか？ そこにはどういう内部過程が含まれるのか？ 現実の発達の過程でそれらはどう表れてくるのか？ こういう問いに答えることを目指して我々は、保育のフィールドでの子どもの活動を定量的に観察するセンサーとそこからの心の

状態を推定する方式を開発している。そして、このようなデータ駆動によるアプローチが、子どもが生まれてから多くの経験を経て独立して思考する主体として行動するようになる過程を客観的に把握できるようになる日を、さらには主体性をより強化する学習環境を科学的にデザインできる日を夢見ている。

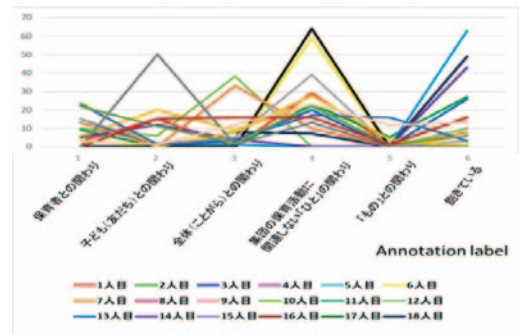


図2 子どもの観察システムと行動分析を通した社会的認知の解明

