

ヒト脳にシンボリックな思考を生み出す 脳アーキテクチャについて

Brain Architecture that Create Symbolic Thinking in Human Brain

大森隆司^{1,2} 宮田真宏²

Takashi Omori^{1,2}, Masahiro Miyata²

¹ 玉川大学工学部

¹School of Engineering, Tamagawa University

² 玉川大学脳科学研究所

² Brain Research Institute, Tamagawa University

Abstract: Brain is a big clump of neurons. Though each of the neurons behaves independently and discretely, the mass of neurons looks to work with continuous value in group. An important feature observed in it is the symbolic thinking that is specific for human brain. The symbolic thinking has features of discrete, logical and conscious that are not seen in current brain understanding and in conventional neural network models, and the features are essential for making human intelligent. In this talk, we overview the history of researches that investigated human brain intelligence, and introduce some researches that may lead to the symbolic thinking study in future.

1. はじめに

ヒトの脳が優れた知能を持つということは自明として、その知能の何がどのように優れているのかという点については、多くの議論がある。実際、ヒトと他の動物の違いはそれほど大きいわけではなく、言語、思考、など一部の能力に限られている [藤田1998]。その中で「思考」について考えると、ヒト以外の動物も思考をしない訳ではなく、ただその質において差があるとされる。ヒトの思考については以下のような記述がある。

「実際に行動として現すことを抑制して、内面的に情報の収集と処理を行う過程。機能的に見て二つの型に分類できる。一つは〈合理的思考〉であって、問題に直面したときにそれにふさわしい解決をめざすという意味で、〈方向づけられた思考〉とも呼ばれる。もう一つは〈自閉的思考〉であって、空想のようにとりとめのない気まぐれな連想によって生じる非現実的思考である。前者は、問題解決のための論理的推論を導く過程であり、概念、判断、推理から成る。(世界大百科事典)」

ヒトにとって特徴的なのはこの「合理的思考」であるとされるが、その詳細はいまだあいまいである。例えば「概念」ひとつとっても、ヒトの言語や推論

に大きく関わる一方で、動物もまたヒトに比べると限定的ではあるが概念を持つことは間違いない。現時点では、生物進化の比較のみから知能の定義・構成要素を明確化することは困難であろう。

一方で、コンピュータで知能を実現しようという動きは1950年代から人工知能(AI)の研究として歴史が始まっている。ここではまず2章で工学的な知能研究とその実現の努力を振り返り、ヒト知能についての工学研究の現在位置を確かめよう。

次いで3章では、シンボリックと言われるヒトの推論はどういう特性を持つのか確認し、ついでそれを生み出すための工学的研究としての認知アーキテクチャについて概観する。そして4章では、これまで著者が試みてきた二つの連想記憶と想起制御に基づく脳アーキテクチャについて紹介し、その背後にある考え方を説明する。そして5章でヒト的なシンボリック思考の実現にむけての一つの可能性としての連想記憶モデルの発展について議論する。

2. ヒト知能のモデル化研究の概観

歴史的には、知的であるということは言葉をあやつる、すなわち論理をあやつること、と考えられてきた。1956年のダートマス会議の時代から、それを象徴する考え方としての記号計算主義、すなわち「我々の思考、特に問題解決に至る有効な思考(アルゴリズム)は記号列の操作として表現される」と

いう考え方が主流であり、それに伴って GPS, 定理証明, エキスパートシステムなど記号の形式操作による知能の実現の研究が多く行われて、すくなくとも限られた範囲では成果を挙げてきた。

一方で、脳のメカニズムのモデル化に基づくニューラルネットによる知能実現の流れも、パーセプトロン[Minsky 1969], アソシアトロン[Nakano 1972][中野 1979], ネオコグニトロン[Fukushima 1980], バックプロパゲーション[Rumelhart 1987], 自己組織化マップ[コホネン 2005]など多くあった。しかし記号処理, ニューラルネットそれぞれ限界があり, 現実的な知能の実現には至らず, 研究は縮小していった。

それを大きく変えたのが 2010 年頃の Deep Learning の出現で, CNN をはじめとする階層ネットワークの高性能化で人に迫る, あるいは分野によっては人を超える性能を示したことは記憶に新しい。これにより, それまでの記号処理的な手法による AI よりはニューラルネット的な手法による AI の方が高機能であるという認識が定着した。

ところが, そのように高機能と言われる階層ネットワークでも, その汎用性という面ではヒトには遠く及ばない。そもそも, ヒトの知能はときに記号処理的であり, ときにニューラルネット的であり, それが時と場合に応じて柔軟に入れ替わり, しかも異なる処理過程が連続して統合されていることに価値がある。そこで注目されたのが認知科学などのヒトの思考の科学である。現在の AI 研究の限界を超えるため, 改めてヒトの思考の理解が求められている。

そこで期待されるのが脳科学である。もともと脳科学は AI と並行して発展してきた経緯もある。その中でも計算論的神経科学は脳の活動を数理として説明しようとしており, 計算機による知能の実現とは相性がよい。これまで多くのモデル研究者が脳研究に参加し, 成果を挙げてきている。

しかし脳モデルは基本的に現象観察からのボトムアップであり, 知能の計算論からは距離がある。ミクロの神経細胞モデルに基づく大規模な脳シミュレーションは可能だが, そこからマクロな計算原理を導くのは容易ではない。例えば, コンピューターの計算原理解明のためにオシロスコープを用いて計算機の配線の信号波形やチップの発熱量を計測することで, 計算中のソフトウェアや OS の原理の解明をしようすることに等しい。現状ではそこまでの解明は困難であるし, そこにボトムアップの脳科学が到達するのはかなり未来のことになる。これは, もともとの理論が無いことによる限界である。

それに対する別のアプローチが, 認知アーキテクチャの構築に基づく構成的研究である。これは知的機能を仮想的に, あるいは実際に作ることで理解を

進めるという方法で, 感覚・知覚・認識・記憶・学習・行動・計画などの多くの機能を再現し統合しようとする, 思考の実現を目指す方法論である。これまで ACT-R から LIDA まで多くの提案があるが, 実際はその部分的な機能の実現あるいはフレームワークの提案にとどまっている。また, その実装の多くは記号処理的な方法である。これは, 機能の実現だけなら脳型である必要はないとの考えであろうが, ルールベースの方法では機能の近似はできても知能の原理にはつながりにくい。

これに対して最近, 自由エネルギー原理や情報統合理論など原理の提案が注目されているが, 提案の抽象度が高いため多様な解釈が可能であり, 現時点では現実の現象に接地しているとは言い難い。

以上, ヒト知能の解明のための研究の流れを概観した。長い期間をかけて知能の理解のための研究は大きく進展してきた。しかし脳の特性を深く考慮した形での思考, 特にシンボリックな思考についてはその原理的なメカニズムが描像されていない。これを解決しないとヒトの脳にせまる思考機械の実現は困難ではないか, と思われる。

そこで本発表では, 脳のニューラルネットの基本構造を強く意識したシンボリックな思考モデルとして, 連想記憶に基づく推論のモデルを説明する。その特徴は, 従来の要素的な知的処理, あるいは認知アーキテクチャに作り込まれた内部手続きを超えた, 明示的な制御機構の導入である。脳とは連想に基づく推論装置であると考え, その振る舞いを直観的推論と論理的推論の間で動的に切り替え, その瞬間のタスクに応じた動作を実現・探索する制御機構の存在を想定している。これにより, ヒト特有のシンボリック思考を説明し, さらに他の動物も行う直観的な思考とも融合することを目指している。

3. シンボリックな思考とは何か

3.1 ヒトの推論

推論についての認知科学的研究によると, ヒトの推論には直観的推論と論理的推論の二種類があるとされ(推論の二重システム仮説), その現象・特性については多くの行動研究・理論研究が蓄積されてきた[服部 2015]。Table 1 はその仮説における二種類の推論の特性の対比である。直観的推論は無意識・速い・直観的といった特徴があり, 論理的推論はその反対に意識的・遅い・抽象的という特徴がある。ヒトはこの二つの推論を適宜切り替えることで, 多様な問題に対して推論を実行可能である。

しかし, ヒトの脳がそれをどうやって実現しているか, そのメカニズムは未解決の重要な謎である。

それが重要な理由は、論理的推論はヒトにしか観察できず言語などの基盤となってヒト知能の本質であるように見えると同時に、論理的な推論には必ず意識が伴うことから、そのメカニズムの解明が「意識」の解明につながると考えられるからである。

Table 1 推論の二重過程と二重システム仮説

直観的推論	論理的推論
作業記憶は不要	作業記憶が必要
無意識的, 自律的	意識的, メンタルシミュレーション
推論が速い	推論が遅い
バイアスに影響されやすい	規範的, 公平
文脈依存	抽象的
確率的, 分散的	論理的, シンボリック
暗黙知(経験的確率)を利用	明示的な知識を利用
推論が浅い	深い推論が可能
進化的に古い	進化的に新しい

近年の脳科学の進歩により、推論には脳の前頭葉、特に前頭前野が関わっていることが知られており、多く生理学的な発見が得られている。例えば坂上はサルの推論課題において前頭葉での意思決定のための推論の瞬間と思われる神経活動を計測している [Pan 2014]。しかし動物はサルでも推論は日常的には行わないようで、多くの推論課題でも過去の判断の記憶に基づく行動決定の脳過程しか観測できていないのが実際のものである。

3.2 認知アーキテクチャの機能原理

現在の多くの認知アーキテクチャの基本デザインは、Paivio の Dual Coding Hypothesis によるところが多い [Paivio 1986]。ヒトの思考には、感覚イメージなどによる直観的な部分と、記述的な知識による論理的な部分の二つが重なっており、それらが相補的に推論を行うという理論である。二つの階層で計算理論が適切に定義されているという点で、それまでのパンデモニウムモデルとは一線を画している [リンゼイ&ノーマン 1984]。この原理は Table 1 に示したヒトの思考の特性をよく表しており、同時代以降の多くの認知アーキテクチャに取り入れられている。

これと同様のデザインは、古典的でありつつ現在も研究が継続されている ACT-R にも含まれている。ACT-R は、知能とは脳内にある多くの機能モジュールの相互作用であると考えられる知的処理のフレームワークである [Anderson 2004]。フレームワークとは、知的システムのデザインの方法論であり、個々のタスクに応じてモデルを実装する際に利用するひな型である。そのため、実現された認知モデルは結局は脳

内の情報処理を表すタスク依存のプログラムにより表現される。その処理の基盤は「宣言的記憶」と「手続き的記憶」で、これらはヒトにより記述されることから ACT-R はシンボリックであると言われる。しかしその意味は、知識や手続きの記述に記号処理的な方法を使うということであり、ヒトの脳のシンボリックな機能のメカニズムを説明している訳ではない。

それに脳科学的あるいは認知科学的な背景、さらにより計算論的なメカニズムを付加した認知アーキテクチャに Ron Sun による CLARION [Helie 2010] や Randy O'Reilly による Leabra がある [O'Reilly 2016]。他にも同様のモデルは提案されており、それらの関係が議論されているほどである [Evans 2009]。

これらの認知アーキテクチャとは一線を画する脳のモデルとして、Baars の意識の Global Workspace Theory (GWT) がある [Baars 1988]。GWT では、脳を巨大な神経回路と考え、その周辺に各感覚モダリティの処理回路が広がり、その先端に感覚器からの入力が入ってくると考える (Fig. 1)。そこに外部から感覚入力が入ってきたとき、入力部からわずかな範囲にのみその興奮がひろがったときはそれは意識に登らない (Fig. 1. T3)。その興奮が T3 より広い範囲に広がるが、やはり途中で止まった場合には、意識に

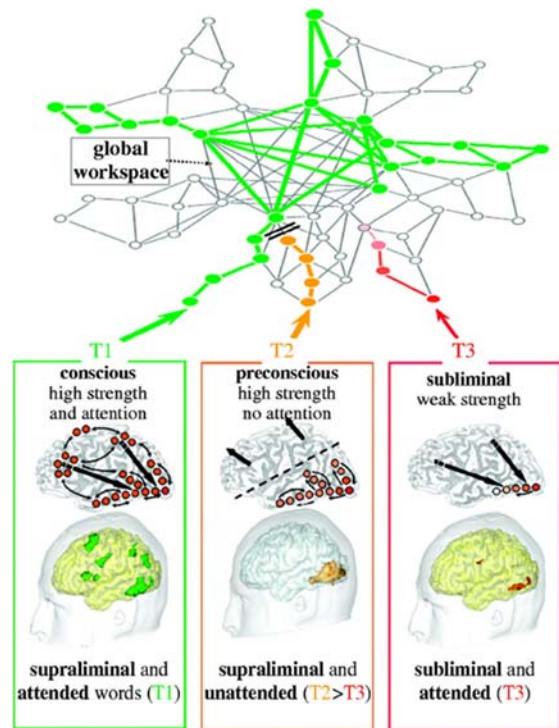


Fig. 1 Global Workspace Theory の概念図
[Dehaene 2011]

は登らないが潜在的に意思決定に影響することもある (Fig. 1. T2)。さらにその興奮が脳内に入ると突然

脳全体に広がり、はじめてヒトの意識に登った状態になる (Fig.1 T1) とする。この意識に関する仮説は、脳科学者の関心を引き、例えば Dehaene などにより検証されている [Dehaene 2011]。

GWT をうけて開発された認知アーキテクチャが Stan Franklin による LIDA である [Franklin 2011][Ramamurthy 2006]。LIDA では、感覚入力・空間記憶・エピソード記憶・運動記憶などの認知機能の要素群の上位に WORKSPACE とよぶ現在の意識状態をあらゆる短期記憶があり、そこから他の要素にその内容を伝搬させる経路がある。LIDA に関しては Franklin による多くの発表があるが、実装はいまだ未完成であるようで、その意味でこれもまたフレームワークに留まっている。Dual Coding Theory の特性も包含して意識につながりうる認知アーキテクチャとして注目すべきものであるが、ここ数年、外部への発表が低下している。

意識に関する理論としては他に統合情報理論がある [トノーニ 2015]。これまでにはない、意識を定量化しようという計算モデルであるが、原理としてなぜ意識が生まれるのかというメカニズムの説明にはなっておらず、現時点では認知アーキテクチャに取り込むのは難しいように感じる。同様に Karl Friston が脳の認識に関する自由エネルギー原理を提唱しており [Friston 2010]、その抽象度の高さから脳の他の機能にも適用できる可能性がある。しかし現状ではその解釈はあいまいであり、これを積極的に取り込んだ認知アーキテクチャはいまのところ存在しない。

3.3 連想記憶にもとづく論理思考のモデル

シンボリックな思考は必然的に意識に関わり、その説明もまた脳全体に関わる活動により説明されるのが望ましい。そこで考えつくのが、連想記憶モデルである。記憶と知能の間には強い関係があり、「知的である」の一つの定義が多くの知識を持つ、すなわち強い記憶力である。推論に使われる知識もまた記憶に保存される。我々は記憶や知識がないと思考もできないのは事実である。

例えば ACT-R や LIDA では各種の記憶を利用している。しかしそれらは計算機的な情報の保持の延長を想定しているように思われる (どちらもフレームワークなので実装については制限していない)。認知アーキテクチャの立場からはヒト脳は記憶をどのように利用しているか、そこからヒントを得たいところである。例えば、神経細胞は記憶を保持するシナプスと演算をする細胞体があったという素子でそれが広く分散しているのが脳であり、記憶と演算を分離・集中させた現在のコンピュータとは大きく異なるアーキテクチャである。どこまでの類似性と

相違を組み合わせるのが、鍵になる。

例えば作業記憶は知能との関係が深く、多くの研究がなされている [Baddley 1974][クラツキー 1984]。神経細胞が他の素子による記憶のモデルとしては連想記憶があり、例えば中野のアソシアトロンから始まる相関行列型の研究の歴史は長い [Nakano 1972][Sompolsky 1986]。連想記憶と推論は近い関係にあり、例えば中野はその著書で連想記憶で推論ができることのひな型を示している [中野 1979]。

ところが 1990 年台をになって相関行列型の連想記憶の研究は急速に数が減ってきて、記憶の機能の実現手法にはこだわらず記号的な手法を用いる認知アーキテクチャが主流となってきた。しかし、脳型の情報処理という視点からは、ニューラルネットによる記憶と連想の実現という意味での相関型の連想記憶の特性は無視できない。そこで望月は 1996 年台に相関型の連想記憶を原型とした推論モデル PATON を提案した [望月 1996]。もとは海馬系と新皮質の連合領野群との関係をモデル化したものだが、脳内の連想過程の手続きを指定する手段として脳内注意の時系列を利用し、論理的推論に見える計算過程を実現している [Omori 1999]。

思考とは問題解決のための道筋を導く過程であり、概念・推論・判断など多くの機能で実現される。これらの要素機能は脳内の個別のモジュールの組み合わせで実現されるというのが現在の脳科学の認識であるが、その組み合わせはどうやって決められているのか、現時点では未解明である。例えば、私たちの日常的な新しい判断行動の素早い獲得は、通常のシナプスの学習による手続き記憶としてはあまりに獲得が速い。これは記憶に近いが、そのような手続き、とくに脳内の処理過程についてのエピソード記憶という概念は現在の記憶の分類には見当たらない。

そこに必要と考えられるのが、脳内過程の制御である。脳における処理の制御にはボトムアップとトップダウンの二種類があると言われる。ボトムアップ制御とは感覚入力に駆動されて自動的に起動される脳内の処理手続きであり、トップダウン制御とは意識的あるいは無意識的に行うタスク処理やその準備のかまへの形成である。このようなかまへの形成は、例えばアセチルコリンなどの神経伝達物質が脳の一部領域に準備的に分配されてそれにより神経回路の活動が活性化されたり伝達特性が変化することで実現できよう。これらの性質を利用して、大森と宮田は連想記憶のトップダウン制御のモデルを提案している。次章はそれについて説明する。

4. 連想記憶によるシンボリック動作

大森らの研究は、連想記憶に外的な制御機能を加えることで、表面的には離散的・論理的に見える推論を実現することに特徴がある。その数理的な基礎は、連想記憶のエネルギー関数である。対象結合をもつ相関関数型の連想記憶は、ある記憶の安定想起状態に至る過程でネットワークのエネルギー最小化のための最急降下の計算を実行し、それが極小値に収束したときに目的とする記憶が想起された状態になっている。そして、脳を一つの巨大な連想記憶と考えるなら、GWT の特定の認識が脳内に伝搬する過程がこの記憶の想起過程に相当するであろう。さらにこの過程は、ある記憶の想起状態から別の記憶の想起状態への状態遷移が起きると考えるなら、外部からは意識の遷移を伴う離散的な推論過程のように見える、というのがこの研究で考えるシンボリックな推論過程の描像である。

ここでは、まずその端緒となった記憶モデル PATON について説明し、ついで宮田により提案された価値を導入した推論モデルについて紹介する。

4-1. PATON

初めて制御を意図的に導入した連想記憶モデルであろう。もとは海馬と新皮質の連合領野の間の接続を想定した構造であり、モデルの構造自体が脳内のマルチモーダルな記憶、すなわち概念の表現となっている(Fig.2)。

その構造は大きく二つの部分に分かれる。一つは、新皮質の連合野を想定したパターン層の複数の領野と海馬を想定したシンボル層と、それらをつなぐコネクション W_{ij}^k である。これらは一体となって連想記憶を構成しているが、シンボル層は多数の細胞のうち一つのみが興奮して、パターン層のマルチモーダルな興奮を連合するハブの役割を果たすデータ構造になっている。この構成は、海馬を中心としたマルチモーダル概念の表現のモデルでもある。

もう一つは、この連想記憶に外部から制御信号を送っているのが Fig.2 の下側にある実行制御部である。そこからは、パターン層の各領野およびシンボル層の各細胞に制御信号が送られており、対象領野および細胞の活性化・抑制・活動の保持・学習といった動作を指定する。制御信号はある瞬間の対象群の動作を規定し、その動作が収束すると次の制御信号が送りだされる。これを繰り返すことで、脳内の複雑な信号の流れを制御する。このモデルが保持する概念記憶と制御信号の列（注意ベクトル）を適切に獲得することで、複雑な推論なども容易に実現できる。すなわち、推論過程に必要な知識のデータ構

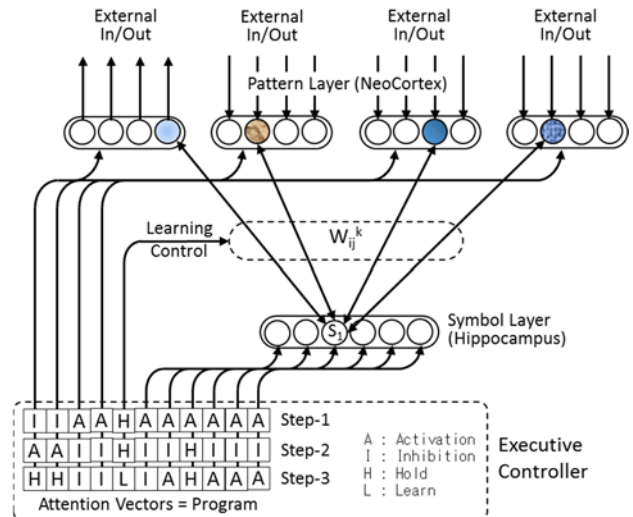


Fig.2 PATON モデルの基本構造。新皮質—海馬の記憶構造モデルに加え、注意ベクトルの列による実行制御システムがある。

造とそれを操作するアルゴリズムを分離したモデルとなっている。ヒト脳の場合は、データとアルゴリズムの完全分離は実際にはなされておらず、人は新しい経験に対して過去の方法を適用するのに苦労することも多い。

このモデルの課題は、概念知識と注意ベクトルの獲得である。論文では概念知識と注意ベクトルは人手により与えられており、その獲得は未解決の問題とされていた。現在の AI の立場から見ると、この部分の構造は LSTM の学習に似ている。しかしこのモデルには報酬に相当する教師信号がないため、そのままでは学習はできない。

4-2. 価値駆動シンボル推論モデル

脳の行動決定において「報酬」または「価値」は欠かせない情報である。そもそもヒトは価値を最大化するように行動するというのが最近の行動経済学の間人理解であるし、機械学習の分野でも強化学習の報酬最大化、バックプロパゲーションの誤差最小化、論理推論の目的状態との一致など、知的に見える行動において何らかの価値を導入して学習や動作を導くことは不可欠の構造であろう。

そこで宮田は、連想記憶のシンボル推論的に見える連想収束過程と価値認識システムを組み合わせ、論理的推論と直感的推論を動的に切り替えながら価値を最大化する状態に到達する経路を探索する推論システムを提案した (Fig.3) [Miyata 2017][宮田 2019]。Fig.3 の上半分は相関行列型の連想記憶が 2 個並列に入力の層と出力の層の間に置かれている。そのうちの一つはエネルギー最小化過程により記憶の離散

的遷移を示す自己想起型の連想記憶，もう一つは2つの異なる記憶の間の遷移関係を直接に記憶して入力から対応する次の記憶を一撃で想起する相互想起型の連想記憶である。その下側に，連想記憶で想起された記憶パターンの中に含まれる複数の記憶項目のうち価値のある記憶を検出してその価値を即座に認識する価値認識層を置いた。

2つの連想記憶はモード切替によりどちらか一方，あるいは両者を混合して使うように設定される。論理推論モードでは，記憶想起の現在状態に対して並列に想起される複数の候補記憶のうち価値の最大となる対象を並列的に探索して遷移する。その探索過程は，連想記憶の収束のダイナミクスに従う数値計算過程となる。また直観的推論では，入力から連想される複数の候補を一撃で発見してそれらの価値を評価できる。これにより，推論関係が単純で価値の競合のない場面では直観的な推論で高速に推論を行い，複数の価値が競合する場面では論理的推論に動作を切り替えて価値最大の連想経路を発見する。その間，価値認識層は想起された記憶パターンまたはその混合パターンに対して価値を瞬間的に認識し，その認識は状態遷移の制御や動作モードの切り替えに利用される。

このような動作の結果として，宮田は複数の報酬が置かれた迷路探索課題において，トラックバックのない効率的な探索が本モデルを用いて実現できることを示している (Fig.4)。詳細は参考文献を参照されたい。本モデルの特徴は連想記憶により並列的な記憶の想起による分岐のある探索の並列化できたことと，価値の導入による分岐先の選択過程を連想の収束のダイナミクスに組み込んだことである。推論は価値を最大化する連想経路の探索であり，価値を即時に認識する機能を並列的な連想の制御に加えることで，離散的な探索で分岐点で頻繁に起きるトラックバックをなくし，目標とする価値のある状態にほぼ一直線に到達する経路を高速に発見できた。

本モデルは連想記憶としては構造が極めて単純であり，いまだ原理モデルの段階にある。これを現実世界に適用するには，PATONのような構造化されたマルチモーダルな記憶システムが必要となる。また生存のための価値は多様であり，複数の価値の間のコンフリクトの解消も必要である。それらはすでに本モデルに導入されており，記憶の収束ダイナミクスにより解決可能である。さらに，そこに価値が加わると，PATONの注意ベクトルの強化学習による獲得が可能となる。その際，注意ベクトルは過去の知識として再利用することで，GA的な探索も可能となり，極めて効率的な脳内手続きの探索につながる可能性がある。

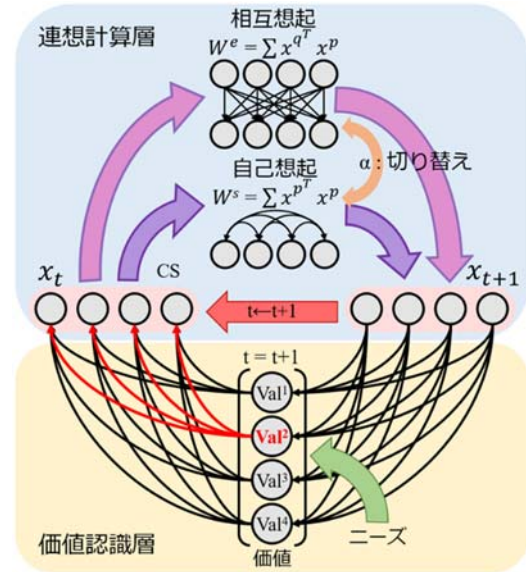


Fig.3 連想記憶と価値認識を組合せた推論モデル

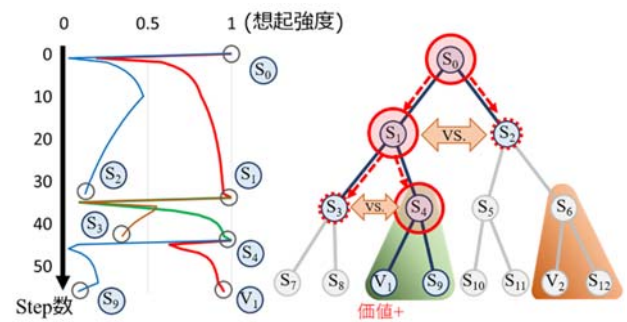


Fig.4 価値に導かれた連想探索による並列 Tree 探索

5. 考察

シンボリック，すなわち離散的な状態遷移が含まれる思考が脳に生まれるメカニズムとしては，現時点ではGWTがもっとも有力な説明であろう。GWTの実現としてのLIDAは，残念ながら現時点で動作している実装はない。

LIDAは各種の記憶に基づき動作するシステムであるため，それを本稿で紹介したPATONのような脳型の記憶構造，さらに価値駆動のシステムなどを導入することとの矛盾は少ない。GWTは連想記憶の記憶の想起状態と極めて相性がよく，あとは複数のモダリティや連想回路の部位を選ぶ注意システムの導入などで，マルチモーダルの大規模な脳型システムでの意識的・無意識的な推論を実現するなら，それは「思考」という言葉にふさわしい複雑さを持つようになると期待する。

相関関数型の連想記憶のもう一つの特徴が，偽記憶の発生である。この分野ではよく知られているこ

とであるが、連想記憶は記憶したパターンを組み合わせた新しいパターンを容易に形成する。これは、これまで連想記憶の欠点と考えられてきたが、統合情報理論の考え方からすると脳型の回路が記憶を覚えるたびに新しい安定した脳状態（記憶の想起場面）を作り、複雑な意識状態を構築するという考え方もできる。また、この偽記憶は記憶した概念の新しい組み合わせという見方もでき、創造性という観点からも検討の余地がある。

一方で、提案した Fig.3 のモデルでは論理的な探索を行う場合に作業記憶がなく、できる探索の範囲が限られていた。論理的には探索には作業記憶が不可欠であり、その容量によってできるアルゴリズムの範囲が決まってくる。その意味で、本モデルを強化していく一つの方向は、作業記憶の付加である。

PATON モデルの課題である手続きの獲得は、基本的には強化学習と遺伝的アルゴリズムの利用を想定できよう。まず当初は試行錯誤が少なく済む短い注意ベクトルを強化学習で獲得し、次いでその知識を過去に獲得された類似場面での手続きと組み合わせることを事前確率付き GA のような方法で行うことで、高速な手続き探索が実現できよう。そのためには、価値がタスクごとに定義されている必要があるが、価値システムはその当面の目的には足りるものと考えられる。

宮田が導入した価値認識システムであるが、ヒトの意思決定は価値の計算に基づくという知見に基づくと同時に、ヒトの情動とは価値を計算するシステムであるという説にも依っている[信原 2017]。進化の中で、情動とはその個体に対する環境の事物の価値を素早く認識するシステムとして個体の生存に有用であり、それがヒト社会では適切とは言えない行動をとることもある。しかし、ヒトにおいても感情はヒトの行動決定の有用な情報源であることは多くの研究で示されている[ヴィンター 2017]。

6. まとめ

以上、知能モデルの研究の流れを概観し、思考につながる研究としての認知アーキテクチャの展開を紹介した。そしてその中でも特に、連想記憶による離散的な状態遷移を用いてシンボリックな推論を実現可能なシステムとして、PATON と価値駆動シンボル推論モデルを紹介した。価値のシステムを含む脳型知能は、結果的にヒトの脳に近い構造を持つように見える (Fig.5)。それは、そもそもが価値システムが先に存在し、そこに付加的な情報処理システムとしての新皮質が進化とともに増大してきたことによる。その価値システムが、ヒトの関心・動機・意欲を作るのであり、その解明なしの汎用人工知能

はヒト的な個性は持たないのではないかと予想する。

本モデルは現時点では概念レベルにとどまっており、現時点で汎用知能の性能を示す段階にはない。しかし Baddeley 以来の中央制御システムの思想に明示的に答えてかつ意識に迫りうる認知アーキテクチャは他になく、今後の検討を続けたい。

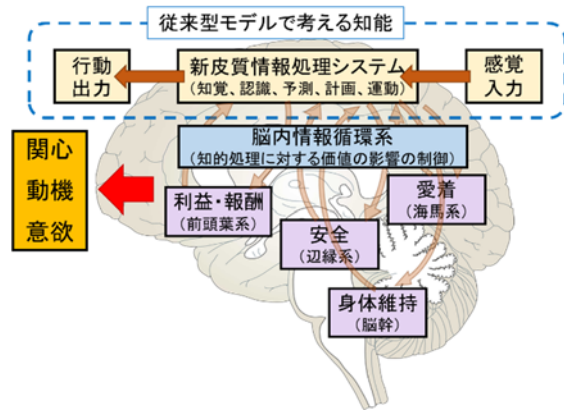


Fig.5 価値システム（情動系）を含む脳型知能の姿

謝辞

本研究の一部は科研費新学術領域「認知的デザイン学」公募研究 15H01662 の助成をうけた。支援に感謝する。

参考文献

- [1] 藤田和生: 比較認知科学への招待—「こころ」の進化学, ナカニシヤ出版, 1998
- [2] Minsky M., Papert S.: Perceptrons: An Introduction to Computational Geometry, MIT Press, 1969
- [3] K. Nakano, Associatron-A Model of Associative Memory, IEEE, vol. SMC-2, pp. 380-388, 1972
- [4] 中野 馨. アソシアトロン—連想記憶のモデルと知的情報処理. 昭晃堂 1979.
- [5] Fukushima, Neocognitron (1980). "A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". Biological Cybernetics 36 (4): 93–202.
- [6] Rumelhart D., McClelland J. : Parallel Distributed Processing, MIT Press, 1987
- [7] コホネン: 自己組織化マップ, シュブリンガーフェアラーク東京, 2005
- [8] 服部雅史, 思考と推論: 理性・判断・意思決定の心理学, 北大路書房, 2015
- [9] Xiaochuan Pan, Hongwei Fan, Kosuke Sawa, Ichiro Tsuda, Minoru Tsukada, Masamichi Sakagami : Reward Inference

- by Primate Prefrontal and Striatal Neurons, *Journal of Neuroscience* 34(4) 1380 - 1396 2014
- [1 0] Paivio, A (1986). Mental representations: a dual coding approach. Oxford. England: Oxford University Press.
- [1 1] リンゼイ&ノーマン著, 中溝, 箱田, 近藤訳, 情報処理心理学入門 II, サイエンス社, 1984
- [1 2] Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 1036–1060
- [1 3] Helie, H. and Sun, R. (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological Review*, 117, 994-1024.
- [1 4] O'Reilly, R.C., Hazy, T.E. & Herd, S.A. (2016). The Leabra Cognitive Architecture: How to Play 20 Principles with Nature and Win!. S. Chipman (Ed) *Oxford Handbook of Cognitive Science*, Oxford: Oxford University Press.
- [1 5] Evans T. et al: How many dual-process theories do we need? One, two, or many?, *Oxford Scholarship Online*, 2009
- [1 6] Baars, Bernard (1988), *A Cognitive Theory of Consciousness*. Cambridge, MA: Cambridge University Press.
- [1 7] Dehaene S., Changeux J.P. : *Experimental and Theoretical Approaches to Conscious Processing*, pp.200-227, *Neuron* 70, 2011
- [1 8] Franklin S., GLOBAL WORKSPACE THEORY, SHANAHAN, AND LIDA, *International Journal of Machine Consciousness*, Vol. 3, No. 2 (2011) 1-11
- [1 9] Ramamurthy U., Baars B., D'Mello S., Franklin S., LIDA: A Working Model of Cognition, *ICCM2006*
- [2 0] ジュリオ・トノーニ、マルチェッロ・マッスィミーニ『意識はいつ生まれるのか 一脳の謎に挑む統合情報理論』花本知子訳、亜紀書房、2015年。ISBN 978-4750514505
- [2 1] Friston. K.(2010). The Free-energy principle : a unified brain theory?, *Nature Reviews Neuroscience*,11(2),127-138
- [2 2] Baddeley, A.D. and Hitch, G.J. (1974) Working memory. In *The Psychology of Learning and Motivation* (Bower, G.A., ed.), pp. 47–89, Academic Press
- [2 3] クラツキー : 記憶と意識の情報処理, サイエンス社, 1986
- [2 4] H. Sompolinsky, Temporal Association in Asymmetric Neural Networks, *Physical review letters*, Vol. 57, No. 22, pp. 2861-2864, 1986
- [2 5] 望月、大森 : PATON: 文脈依存性を表現する動的神経回路網モデル、*日本神経回路学会論文誌*, Vol.3, No.3, 81-89(1996)
- [2 6] Omori T., Mochizuki A., Mizutani K., Nishizaki M.: Emergence of Symbolic Behavior from Brain Like Memory with Dynamic Attention, *Neural Networks*, Vol.12, No.7-8, p1157-1172, 1999
- [2 7] M. Miyata, T. Omori: Modeling emotion and inference as a value calculation system, *BICA2017*, vol. 123, pp. 295-301, 2017
- [2 8] 宮田, 大森 : 価値に駆動された連想記憶に基づく人の推論過程の統合モデルの提案, *知能と情報(日本知能情報フレンジ学会誌)*, Vol.31, No.3, pp.712-721, 2019
- [2 9] 信原幸弘 : 情動の哲学入門 価値・道徳・生きる意味, 勁草書房, 2017
- [3 0] エヤル・ヴィンター : 愛と怒りの行動経済学ー貸し越し人は感情で決めるー, 早川書房, 2017