

3. 人工知能のホットピック

1 汎用性の創発を脳に学ぶために

山川 宏 ((株)ドワンゴ ドワンゴ人工知能研究所)

表現と推論

深層学習の技術的進展が近年の第三次人工知能ブームを牽引してきた。この流れを引き継いで人工知能を今後発展させていく方向性として、汎用人工知能や脳型人工知能がある。いずれも以前から存在する研究アプローチではあるが、深層学習技術を取り込むことで大きく前進し、その間の関係性も様変わりした。

本稿では、人工知能技術の基本的な構成要素である、表現と推論という2つの側面を軸に私なりの視点から本動向を俯瞰する形での解説を試みたい(図-1参照)。

➡ 記号表現と分散表現

人工知能において知識の記述に用いられる表現は、大きく記号表現と分散表現に分けられる。

- 記号表現：局所的な値に特定の意味(セマンティクス)を対応付けた表現。自然言語における単語等に相当する。コミュニケーションやプログラミングに利用できる。
- 分散表現：多次元ベクトル等における値の分布を用いて概念を表現する。通常は外部からの意味付けを行わないため、コミュニケーションやプログラミングへの利用は困難だが、多次元空間中で概念間の距離や関係等を用いた処理は可能である。

➡ 帰納推論と演繹推論

人工知能に用いられる論理的推論は、主には帰納推論と演繹推論であり(本稿ではアブダクションは議論しない)<sup>1)</sup>、人工知能の観点からは以下のように捉えられる。

- 帰納推論：データから知識を獲得する推論。機械学習全般を含む。
- 演繹推論：知識を組み合わせて活用する推論。データがない場合や少ない場合にも利用できる。

表現と推論の両軸から人工知能技術を俯瞰しつつ(図-1参照)、各々について説明する。

➡ 知識を活用する演繹推論

古典的人工知能の流れとして記号表現を用いた演

	記号表現	分散表現
演繹推論 (知識活用)	古典的人工知能 (1950年代~) 【大人の知能】	分散表現演算 Word2Vec等 (2010年代~)
帰納推論 (知識獲得)	パラメトリック機 械学習、 帰納論理プログラミ ング等 (1960年代~)	ノンパラメトリック 機械学習、 表現学習機能 (1990年代~) 【子供の知能】

図-1 人工知能技術における表現と推論  
図中で①と②は分散表現による知識に対して演繹推論を行う2つの方法を示している。

繹推論に関する研究が1950年代から始まっている。これは小学生以上で発達する計画や論理思考といった、いわゆる大人の知能である。

本稿では、分散表現に対して演繹推論を行う技術を分散表現演算と呼ぶことにする。近年、この種の研究が盛んである。たとえばWord2Vecではように分散表現上で類推を行う機能などが実現されている。

➡ 帰納推論による知識獲得と事前知識

機械学習の学習器は、データから知識を獲得する帰納推論のための装置である。一般的に学習器はデータを利用して数学的なモデルに含まれるパラメータを調整することで知識を獲得する。よって学習器が有する知識は、設計者が事前にモデルに組み込んだ知識と、データから獲得した知識であり、定性的には、それら知識の総和により性能の上限が規定される。

学習器が対象とするタスク範囲が狭ければ、より多くの事前知識をモデルに組み込み得るし、そこに含まれるデータの性質も揃うので、良い性能を得やすい。逆にタスクの範囲が広がると事前に共通的に設計できる知識が少なくなり、性能を上げることが難しくなる。

機械学習が発展してきた歴史を振り返ると、基本的には、計算量とデータの増大に伴って、設計者が組

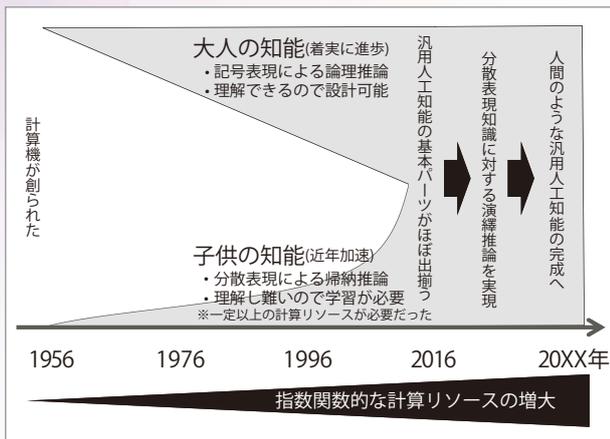


図-2 大人の知能と子供の知能の発展と合流  
2015年頃から2つの知能が合流し始めたことで、汎用人工知能研究が本格化しはじめた。

み込んだ知識よりもデータから獲得された知識へのウェイトが増している。

初期の機械学習の研究は、線形回帰モデルのように明示的に意味付けし得るパラメータの調整から始まっている。当然ながらこうした手法では、設計時の想定を超えた表現は獲得されない。

設計時にパラメータの意味を決定しなければ、潜在的には学習で獲得した表現の一部分については外部からの観測で意味付けし得る可能性がある。その後発展した人工ニューラルネットワーク(ANN)などのノンパラメトリック機械学習がこれにあたるが、深層学習の発展を経てそれは現実のものとなった。これは乳児が認知発達を通じて暗黙的に獲得する知能の性質に合致することから、ここではこれを子供の知能と呼ぶことにする。

近年の深層学習の成功は<sup>2)</sup>、基本的にはANN技術に対して高速計算と大規模データが適用できる環境が整った結果である。重要なことは「十分にデータを得られるタスクの範囲内であれば、応用価値のある人間並みの性能を持つ帰納推論が可能になった」という点である。つまり計算パワーの増大により、最近ようやく子供の知能が実現されたのである。こうして、図-2に示すように大人の人工知能と子供の人工知能が出揃うことで、その統合を通じた人間のような人工知能の実現に向けた動きが本格化し始めているのである。

## 人型人工知能と脳型人工知能

人工知能の研究開発を進めるにあたり、目標として

人の振舞いもしくは人の脳に似せて作るか否かは重要な判断軸である。

そこでまず人工知能一般、人型人工知能、脳型人工知能を、研究目標の面から整理すると以下のようになる。

- 人工知能一般の目標：産業や科学技術開発等の応用を志向する、何らかの意味でより高度な知的能力の実現
- 人型人工知能に固有の目標：人のように振る舞う知能の実現
- 脳型人工知能に固有の目標：脳の機能的な理解に基づく医療貢献、さらにはマインドアップロード等の実現

人工知能に期待される社会貢献や応用シーンでは人や脳に固有な知能は必ずしも必要とされない。一方で、人工知能研究は常に人の知能との比較をしてきた。こうした関心が高いのは、あるタスクにおいて人工知能の知能が人間を上回れば、それにかかわる職業は人工知能に代替される可能性が高いからであろう。

さらに脳型人工知能では脳メカニズムから医療に貢献する側面なども研究目標に含まれ得る。

人工知能一般を目標とした場合には、人型人工知能や脳型人工知能を目指すことは到達手段になる。同様に人型人工知能を目標とした場合には脳型人工知能は到達手段になる。脳型人工知能において脳全体を扱うのであれば、人型人工知能は1つの目標となり得る。

## 汎用人工知能

### → 汎用性という技術目標

汎用人工知能は、現在実用化されている特化型人工知能に対置して現れた概念であり、2006年頃にBen Goerzel氏により提唱された。大雑把に言えば多種多様な課題に対して問題解決を行える人工知能を構築しようとする試みである。逆に現在実用的な特化型人工知能は、対象とするタスク(たとえば、囲碁、自動運転など)に応じた事前知識が潤沢に組み込まれている。

一方、汎用的な人工知能に期待される知能として以下の3つがあるだろう。

- 人間が備えている知能
- 現状の人工知能で実現されていない知能

### 3 人工知能のホットピック… 1 汎用性の創発を脳に学ぶために

- 性能評価できる知能 (実は難しいのだが)

汎用人工知能の性能評価に対する指標は主に2種類ある。先の人工知能一般を目標とする立場では「経験からの学習を通じて、さまざまな問題に対する多角的な解決能力を獲得できる知能」となる。対して人型人工知能の立場では、人と同程度に多種多様な知的能力を発揮できるという側面から評価される。

逆に「汎用人工知能は何ではないか」を指摘するならば、以下のように言えるだろう。

- 単に特化型人工知能の寄せ集めではない
- 最初から何でもできる知能ではない
- タブラ・ラサ (白紙) から学習するのではない
- 意識の有無を考慮しない (評価が困難)

#### ➔ 深層学習の後押しで本格化する汎用人工知能研究

個別のタスクに着目すれば、大量データが得られる状況であれば ANN 等の帰納推論が人を超えた知的能力を実現できることも多い。

こうした背景から、最近の汎用人工知能の国際会議などにおいて議論される人工知能の基本課題は個別のタスクを離れ、主に以下のような側面が着目されている。

- 汎用性
- 演繹推論 (計画などの大人の知能)
- 少数データへの対応
- 現実的な時間内での問題解決

実はこの4課題は密に関係する。知能の汎用化を目指してタスクの適用範囲を広げていけば、相対的にデータ不足となる。しかし眼前のタスクに対して新たにデータを収集するほどの時間的余裕がなければデータ不足は解消できない。すともはや帰納推論には頼れないため、演繹推論を導入する必要が生ずる。つまり今後の人工知能研究開発における基本的課題として、汎用性という側面はより鮮鋭化したと言えそうである。

しかしながら図-1に示したように、ANN等の帰納推論では分散表現として知識が獲得されるが、伝統的

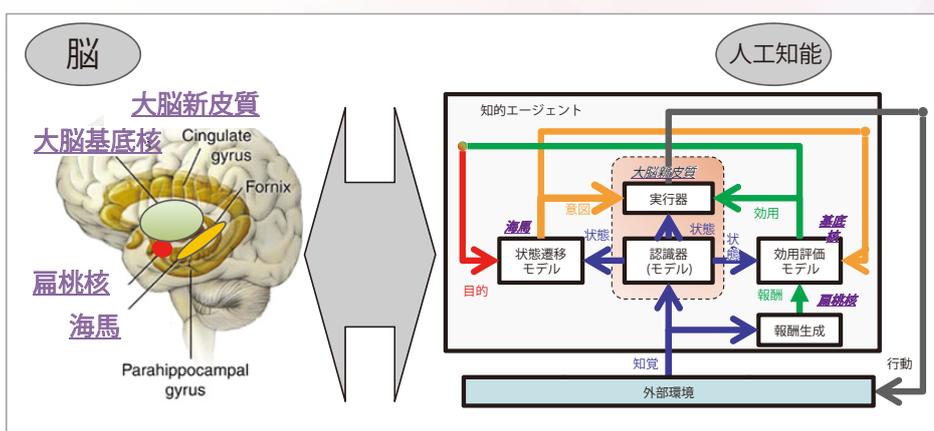


図-3 脳に対応付けられた人工知能

な演繹推論は記号表現を用いて行われる。つまり「分散表現知識に対する演繹推論」の実現という課題が生ずる。

#### 脳型人工知能：脳を参照して汎用性に迫る

脳型人工知能の基本的な目的は人工知能を作ることにある。より確立した近隣分野として計算論的神経科学<sup>3)</sup>の分野が存在するが、こちらはむしろ理解するための計算論であり目的が異なるが、互いに参考になる学術領域である。

#### ➔ 脳型人工知能の実現性が高まった

現存する汎用人工知能は脳以外にはなく、第一義的にはこれを真似て汎用人工知能を構築する脳型人工知能のアプローチは早道かつ自然に思える。ANNは脳の神経回路を模したものであるし、深層学習は脳の視覚野の階層構造を参考にしている。また大脳基底核の機能は強化学習と対応付けることができる。全般的に見ても図-3に示すように脳に対応付ける形で、人工知能の基本的な仕組みを対応させることが可能である<sup>4)</sup>。

これまでは脳型人工知能を推進するにあたり2つの大きな課題があったが、解決されつつある。

1つ目に、汎用人工知能の特性を生み出す上で重要な役割を担う大脳新皮質に対して、工学的に有用な情報処理を行える計算モデルが存在しなかったことである。それをある程度模倣できる形で深層学習が現れた。ここで新皮質は、その全領域に渡りほぼ共通の6層構造を持つ局所回路で構成されるが、入出力される情報に応じて異なる機能を実現する<sup>5)</sup>。これが学習によって多様な機能を獲得する汎用人工知能の特性に対応する。

2つ目は、神経科学の知見を人工知能に応用するこ

とが必ずしも容易ではなかったことである。その大きな原因は、従来の神経科学知見は脳全体のマクロな振舞いと、神経細胞数個のミクロな振舞いへの理解が切り離された形で進展するというマイクロ-マクロギャップが存在したためである。しかし現在は、光遺伝学の進歩により、動物実験ではある部位における1,000個規模のニューロン活動を同時計測が可能になり、その神経活動を制御できるようになった。また脳全体の静的なネットワーク構造や、局所的な詳細なネットワーク構造がコネクトーム研究で明らかにされつつある<sup>6)</sup>。さらに脳内で離れた複数領域の活動を同時測定し、その関係を明らかにするさまざまな研究も進んでいる。つまり近年、神経科学の研究は階層的なネットワークの動的な理解に立ち入りつつあり、ようやく人工知能の設計にとって有用な段階となってきた。

#### ⇒ 汎用人工知能開発において脳を参照する意義

ストレートに脳の仕組みを同定して、それを計算機上に実装することで脳型人工知能が作れるなら分かりやすい。しかし、そこまでの脳の理解が進むにはいまだ長い年月を要しそうである。そこで現状の脳型人工知能は、脳をあくまで参照として用いる、というスタンスをとることになるが、脳型人工知能のアプローチには、2種類の知見を得ることができるというメリットがある。

#### 【A】 設計のガイドに関する知見

- 現状の人工知能で実現されていない未解決な計算機能についてのヒント（暫定モデル）
- 目的が多岐にわたる汎用人工知能は、目的に応じた機能分割による設計が困難であることに対処するためのアイデア

#### 【B】 分散共同開発に関する知見

- 開発初期段階から脳の機能と構造に合致するように実装を制約することによる、後々に技術統合の問題が生じるリスクの回避
- 周辺の科学的知見（認知科学・神経科学等）を集約する足場の構築
- 汎用人工知能に必要な機能の抜けの確認
- チーム内で「脳に近い実装を優先する」という価値観を共有することによる、大規模分散開発の発散の防止

#### ⇒ 分散表現に対する演繹推論は脳でいかに創発しているか

今後の汎用人工知能にとって課題となり得る分散

表現知識に対する演繹推論について考える。すると、解決法は以下のいずれか、もしくはその両方となるであろう（図-1参照）。

- ① 記号接地アプローチ：分散表現知識を記号的知識に対応付けた上で、記号により演繹推論等を行う
- ② 分散表現演算アプローチ：分散表現知識に対して直接に演繹推論等を行う

上記の機能が脳においてどのように実現し得るか想像してみたい。

1つ目にかかわる記号接地は言語機能にかかわる、長年にわたる人工知能の未解決課題である。これは大脳新皮質上において話す機能を担う領野（ブローカ野）や聞く機能を担う領野（ウェルニッケ野）との関連が深く、機能的なモデル化なども進んでいる。

2つ目の分散表現演算については、脳に対応付け議論を行った研究はまだ目にしたことはない。また、多くの動物はほぼ演繹推論を行えない。よって人間の脳においては分散表現演算を実行していたとしても、ある程度は記号表現と連携している可能性が高いと想像している。

いずれにしても多数の脳の領野が神経ネットワークを通じて連携することで高度な機能を創発しているものと思われる。そして今後さらに進展が続くであろう神経科学分野から、こうした汎用人工知能実現に向けた課題にブレイクスルーを与えるヒントが生まれると期待している。

#### 参考文献

- 1) 米盛裕二：アブダクション—仮説と発見の論理，勁草書（2007/9/20）。
- 2) LeCun, Y., Bengio, Y. and Hinton, G.: Deep Learning, Nature 521, pp.436-444 (28 May 2015) .
- 3) 銅谷賢治：計算神経科学への招待，脳の学習機構の理解を目指して，臨時別冊数理科学，2007年12月号。
- 4) 全脳アーキテクチャとは，<http://wba-initiative.org/wba/>
- 5) 山川 宏，市瀬龍太郎，井上智洋：汎用人工知能が技術的特異点を巻き起こす，電子情報通信学会誌，Vol.98, No.3, pp.238-243 (2014)。
- 6) セバスチャン・スン：コネクトーム：脳の配線はどのように「わたし」をつくり出すのか，草思社 (2015/11/18)。

(2016年8月5日受付)

山川 宏 hiroschi\_yamakawa@dwango.co.jp

1965年埼玉県生まれ。NPO法人全脳アーキテクチャ・イニシアティブ代表。人工知能学会 副編集委員長。電気通信大学大学院情報システム学研究科客員教授。玉川大学脳科学研究所 特別研究員。人工知能学会汎用人工知能研究会主査。産総研人工知能研究センター客員研究員。専門は、人工知能、特に、認知アーキテクチャ、概念獲得、ニューロコンピューティング、意見集約技術など。