

dwango

DWAL-TR-2015-003

Jul. 21 2015

第10回
全脳アーキテクチャ勉強会
「全脳アーキテクチャのいま」
報告書

中島 隆夫*1 土屋 敏明*2 中村 政義*3 山川 宏*3
Takao Nakajima Toshiaki Tsuchiya Masayoshi Nakamura Hiroshi Yamakawa

*1 三井情報株式会社
MITSUI KNOWLEDGE INDUSTRY CO., LTD.

*2 株式会社DHCシアター
DHC Theater Co., Ltd.

*3 株式会社ドワンゴ 人工知能研究所
DWANGO Co., Ltd. Artificial Intelligence Laboratory

 ドワンゴ
人工知能研究所

DWANGO ARTIFICIAL INTELLIGENCE LABORATORY

本報告書の概要

全脳アーキテクチャ勉強会は、人間のように柔軟汎用な人工知能の実現に興味のある研究者、脳に興味のあるエンジニア、関連分野(神経科学、認知科学等)の研究者間での交流を促進し、全脳アーキテクチャを実現するために発足された。開催 10 回目を迎えた今回の勉強会では「全脳アーキテクチャのいま」をテーマとして本勉強会の原点に立ち返りつつ、約 1 年半の歩みを主要関係者 6 名が講演した。本報告書では各講演の内容を報告する。

0. ご挨拶（日本電気株式会社 執行役員 江村克己様）

本勉強会に会場を提供して下さった日本電気株式会社の江村様よりご挨拶を頂戴した。

《技術の来し方と行く末》

NEC では世界中の人間が言語の壁を超えて自由に話せる社会を目指して、自動通訳の研究開発を長年にわたって行っている。その間にも計算機の性能はいわゆる「ムーアの法則」に従って向上してきた。しかしその法則にも近年陰りが見えてきたとする指摘がある中、またシンギュラリティというパラダイムへの転換点が提唱されるようになった中、技術はどのような方向に進んでいくのだろうか？そのことを考えるにつけて本勉強会の活動は興味深く、学んでいきたい。

《イノベーションの可能性》

異分野の人間が集まることがイノベーションの起きる条件の一つである。NEC は電子情報系の人間が多く、勢い発想が一面的になりがちだが、本勉強会のような多様な人材との交流の中から新しいアイデアが生まれることを期待している。

オープンイノベーションという意味では、NEC 北米研究所が Deep Learning の OSS 「Torch」を開発しており本勉強会とも技術的接点があるので、このような場を借りて様々な観点から議論を深めていきたい。

1. オープニング：全脳アーキテクチャの全体像（東京大学 准教授 松尾豊氏）
2. 講演：人工知能の難問と表現学習（同）

※ 講演資料は <http://www.sig-agi.org/wba/10> 上に公開

2013年の発足以来の“全脳アーキテクチャ小史”を紐解きつつ、松尾先生のご専門である Deep Learning を切り口にして今後の展望が解題された。

《全脳アーキテクチャ勉強会の設立からこれまでの振り返り》

2013年の夏、松尾氏・一杉氏・山川氏が脳機能の実現する日は近いということで意気投合したが、その実現のためには何が必要か？という具体論がクリアではないので勉強会の形で立ち上げたのが“全脳アーキテクチャ勉強会”。マニアックな内容にもかかわらず、毎回 200 名ほどの参加があった。

運営していく過程で若手の会の発足や、人工知能研究拠点の相次ぐ設立（ドワンゴ人工知能研究所、リクルート人工知能研究所、産業技術総合研究所・人工知能研究センター等）があり、人工知能研究の環境はここ 2 年で大きく変わってきている。

《全脳アーキテクチャの基本コンセプト》

全脳アーキテクチャが目指すのは「人レベルの、そして人を超える人工知能」である。作業仮説として採用しているのは

脳はそれぞれよく定義された機能を持つ機械学習器が一定のやり方で組み合わせられる事で機能を実現しており、 それを真似て人工的に構成された機械学習器を組み合わせる事で人間並みかそれ以上の能力を持つ汎用の知能機械を構築可能である
--

ということである。

この仮説を是として構成論的なアプローチで脳を理解し、またその実装が汎用知能機械となる。それが全脳アーキテクチャの目指すところである。

特に主要な構成要素とみなされている Deep Learning の近年の目覚ましい発展については、各種メディア等を通して非専門家の知るところともなっている。日本における Deep Learning の一般向け紹介は日経ビジネス 2013 年 4 月が初出と思われるが、世間一般における人工知能の認識という意味でも、ここ 2 年で大きく変わってきている。

《人工知能全般が直面している困難と Deep Learning》

人工知能 60 年の研究の歴史において、「知」を実現したいと考えつつ、未だに実現していないのはなぜか。よく議論になる困難に以下のようなものがある。

- ・機械学習における特徴量の設計（汎用理論の不在）
- ・フレーム問題

・シンボルグラウンディング問題（記号接地問題、現象の意味的理解・ラベル付けの困難性）

一見異なる問題のように見えるが、その根底にあるのは「筋のいい特徴量を抽出して概念構造として構成すること」を人工知能に担わせられていない、言い換えればコンピュータが解くべき問題の“表現”は、全て人が行っているところにある。裏を返せば、真の人工知能と呼べるようになるには、人間が何も考えずに機械にデータを投入すると自動的に“表現”を創り出せるような仕組みにする必要があり、Deep Learningにはその可能性がある。

Deep Learning は画像認識など比較的“低次の”問題に対して非常に大きな成果を挙げて注目されているが、今後は音声や動画などの時間方向を含む認識、マルチモーダルな抽象化に適用範囲が広がっていくだろう。真価はその先の、表現獲得・意味理解という範疇でこそ発揮されると考えている。

具体的には、名詞のラベル付けだけをしているレベルから行為の抽象化としての動詞・形容詞・副詞の概念を扱えるようになる → 「行為とその帰結」を取り出せるようになる → 文に相当するような高度な言語理解が可能になる → 言語データを大量に入力して学習することで知識獲得、さらには社会の集合的知識獲得という意味での高次社会予測が実現 というステップが考えられる。

《Deep Learning の最新動向》

足元の研究動向という意味でもその進化は加速している。Deep Learning における layer 数は高々8程度がこれまでの相場であったが、最新の研究では20前後を達成、さらに100層レベルの実現に向けた議論が行われている。

画像認識においてはもはや人間の能力を超えたというのが専らの見方だが、ここ最近では認識率の向上にしのぎを削るというよりも、リアルタイム性の向上など実際のアプリケーションとして可用性を向上させる取り組みに軸足が移ってきている。

また Deep Learning には理論がない、という批判がよく見受けられたが、理論研究（ブラックボックスの中で起きていることの分析）も盛んになっていきている。Deep Learning の高度化（高次機能化）に重要な役割を果たすかもしれない。

DeepMind 社が開発した人工知能「deep Q-network (DQN) ¹」がゲームの攻略方法を学習し人間以上のパフォーマンスを出すなど様々なタスクへの適用が試みられるようになってきており、今まさに汎用化への端緒についたと言ってもいいだろう。

¹ レポーター註： 原著論文は”Playing Atari with Deep Reinforcement Learning”
<https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf>

《人工知能が社会に与えるインパクト》

人工知能技術の発展に伴ってそれが社会の中でどのように使われ、インパクトを与えるかという視点も重要である。現在の人工知能の社会における立ち位置は、1995年前後のインターネットの状況と似ている。キープレーヤー・プラットフォームはこれから生成・集合・離散を繰り返しながら産業構造が大きく変わっていくであろう。

適用のされ方という意味でも、いわゆる異常検知のような単純なレベルから状況に応じた柔軟な判断、さらには仮説生成・実験計画の設定・試行サイクルの高速化というよりメタな領域へ拡張されていくと思われる。

そこではまったく新しい産業構造が立ち現れているはずだ。

《そのような状況下での日本の行く末》

日本では第五世代コンピュータ²の取り組みによるおかげで、人工知能の研究者を多く輩出してきた。人数でアメリカに比肩し得るのは日本くらいのものである。また世間一般の関心も非常に高い。インターネットの世界ではアメリカに完敗したことがトラウマになっているが、

- ・インターネット：次々と新しい試みをすることでサービス化の芽を探る
→ アメリカ向き
- ・人工知能：本質の地道な積み上げで高度化していく、「賢さと真面目さ」が必要
→ 日本向き

という特性があると見ており、日本には大きな希望がある。Deep Learning に関して言えば歴史も浅く、人材が集中すればすぐに世界のトップに踊り出することも十分可能だ。それが実現するためには各人が知的好奇心から参加できて、既存の社会的枠組みに縛られず自由闊達に議論できる場が必要であり、全脳アーキテクチャ勉強会がその一端を担ってほしい。

主な参考文献（レポーターによる追記）：

- K. Fukushima (1980) “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*”, 36(4), 93-202
... Deep Learning 技術の源流としてのネオコグニトロン
- McCarthy and Hayes (1969) “Some Philosophical Problems from the Standpoint of AI”, *Machine Intelligence* vol. 4, pp.463-502

² レポーター註： 経済産業省所管の審議会における次の資料を参照 「イノベーション政策としての国家プロジェクトの 評価システムの在り方について」

http://www.meti.go.jp/policy/tech_evaluation/c00/C0000000H23/120113_hyouka41/41hyoukasyoui-8.pdf

... フレーム問題の一次資料

- J・マッカーシー, P・J・ヘイズ, 松原仁著 三浦謙訳 (1990) 「人工知能になぜ哲学が必要かーフレーム問題の発端と展開」 哲学書房

... フレーム問題の概観

- Stevan Harnad (1990). "The Symbol Grounding Problem", *Physica D*, 42, 335-346

... シンボルグラウンディング問題の一次資料

3. 講演：全脳アーキテクチャと大脳皮質モデル BESOM の実用化研究の構想（産業技術総合研究所 人工知能研究センター脳型人工知能研究チーム 一杉裕志氏）

※ 講演資料は <http://www.sig-agi.org/wba/10> 上に公開

最初に全脳アーキテクチャの一構成要素“大脳皮質モデル”に対応するシステム「BESOM」について、そのコンセプトと開発状況が示された。今後 2 年間程度は BESOM に注力し、5～10 年後の実用化を目指すとのことであった。

一杉氏の所属する産業技術総合研究所ではこの 5 月 1 日に「人工知能研究センター」が設立され、一杉氏は「脳型人工知能研究チーム」において外部研究者とも連携しながら BESOM の改良を進めていく計画とのことである。

最後に BESOM もその基礎を置くところのベイジアンネットワークによる汎用情報処理装置についてその構想が説明された。

《全脳アーキテクチャ勉強会の収穫》

- 会を運営する側が勉強になる。
- 脳全体のアーキテクチャの「模倣」に大きな障害が無さそうだとの感触を得た。
- 特に大きな収穫としては、海馬と SLAM の類似点と相違点の理解が進んだ。
- 思考、モデルベース強化学習、周辺化の近似の見通しが良くなった。
- 時系列学習（ESN、LSTM）の研究動向

《BESOM の概要と機能》

BESOM は端的には Deep Learning と同じ構造を持った「制限付きベイジアンネット」と捉えることができる。そのネットワーク構造は feed forward 型ニューラルネットワークと同相で、教師なし学習により特徴抽出するという意味では Deep Learning と同じ機能をもつ。

BESOM は大脳新皮質における神経ネットワークの構成と形態学的に相同であることが確かめられており、そうであるならば機能面でも同じではないのか？ということが作業仮説になっている。

しかし動作させるのは容易ではなく、2007 年に隠れノード数 1 からスタートして、動作プロセスの改良（スパース符号化等）と計算量の削減などを通してステップ・バイ・ステップで規模と機能を拡張してきた。ベンチマークという意味では、最近になって MNIST の手書き文字が認識できるレベルまで到達してきている。

《神経科学現象を再現することによる BESOM の有効性検証》

BESOM 自体の精度向上や並列化などは当然中心的課題であるが、神経科学現象を再現することで大脳皮質の汎用的プラットフォームとしての BESOM の有望性を示すことができる。その具体的ターゲットとしているのが視覚野モデルおよび言語野モデル

で、いずれも人間がするような“深い理解”に迫ることが目標である。

前者の視覚野モデルは低次脳機能として脳科学の世界でも多くの知見がある分野であって、そのアーキテクチャを BESOM で模倣することで同様の機能が得られるかが焦点となる。例えば、行動経済学者のカーネマンらが主張している“オブジェクトファイル理論”³（視覚画像から抽出された特徴の、機能レベルでの取り扱われ方に関する理論）などをアーキテクチャの候補として実証実験を進めていく。

後者の言語野モデルについては、実験動物であるところのサルには言語がない、という理由から神経学的知見が非常に限られている。しかし脳科学の知見こそ比較的乏しいが計算言語学・計算機科学（人工言語）研究の膨大な蓄積がある。言語の独特の構造は脳の構造に対応しているのではないかと考えており、それらの分野が主張する言語処理機能構造のアーキテクチャを模倣することで言語野の機能を理解する、という態度で臨みたい。例えばチャートパーサ⁴を実装するようなアイデアである。

《“汎用情報処理装置”への展望》

さらに“汎用情報処理装置”のレベルまで到達することを目標にすると、脳の高次機能を再現するのに最低限必要な諸サブ機能（変数バインディング、可変長構造、再帰的構造...）を実現する必要がある。昨年 Deep Mind 社が発表した“Neural Turing Machine”が一つの可能性を示している。彼らはニューラルネットをあくまで情報変換器として捉えてメモリについては外部化しているが、メモリも embed した形でのベイジアンネットワーク（BESOM）として実装したいと考えている。

先行研究として、1999年に大森らにより提唱された“PATON”⁵というアイデアに注目している。

PATON においては、個別領野の表現を“状況に応じて”変換する仕組み（“注意システム”と呼称）を用意して、その出力を統合知覚に対応づけている。それは CPU アーキテクチャにおける“マイクロプログラム”と同様なアイデアであり、ベイジアンネ

³ レポーター註： オブジェクトファイル理論の原著論文
Kahneman, D., Treisman, A., & Gibbs, B. J. (1992) “The reviewing of object files: Object-specific integration of information”, *Cognitive Psychology*, 24, 175-219
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.397.1616&rep=rep1&type=pdf>
概説については電子情報学会「知識の森」所収の次の資料を参照
http://27.34.144.197/files/S3/S3gun_02hen_07.pdf#page=6

⁴ レポーター註： Chart Parsing 法については例えば次の資料を参照
<http://ocw.osaka-u.ac.jp/language-and-culture-jp/language-and-information-science-jp/lecturenoteno-8.pdf>

⁵ レポーター註： PANTON の原著論文
大森隆司, 水谷健太郎, 下斗米貴之 (1999) 「PANTON: 脳におけるシンボル処理の発現機構のモデル化」, 日本神経回路学会 99 年度大会
<http://kushsharo.complex.eng.hokudai.ac.jp/~omori/Papers/PATON.jnns99omori.pdf>

ネットワークで実装可能である。

問題は、そのマイクロコードのロジックを如何にして“状況に応じて”変更するか（自律的に！）ということであるが、評価関数の EM アルゴリズムによる最尤推定で学習可能であって、つまりこれも実現可能である。

以上の方向性での汎用化の試みについても今後追求していきたい。

《質疑応答》

Q: 脳の動作原理解明にあたっては、脳内情報処理におけるトップダウンとボトムアップの融合の仕方、嘘をつくという行為の理解などが切り口になるのではないか。

A: 脳を機械学習アルゴリズムとみなすと、そのような現象の見通しが非常に良くなる。例えばウソの本質は報酬期待の最大化であって、トップダウンとボトムアップの組み合わせ方も含めて“評価関数の最大化を目的とした時のパラメタ最適化”として記述できるという意味で。

主な参考文献（レポーターによる追記）：

- 一杉裕志（2015）「全脳アーキテクチャと大脳皮質モデル BESOM の実用化研究の構想」
<https://staff.aist.go.jp/y-ichisugi/brain-archi/20150514wba.pdf>
... BESOM 開発者による BESOM 解説資料
- 一杉裕志（2009～）「BESOM ブログ」
<http://besom1.blog85.fc2.com/>
... BESOM 開発者による BESOM に関するブログ記事
- 中原裕之（2014）「脳の計算理論：強化学習と価値に基づく意思決定」, 理化学研究所 脳科学総合研究センター 理論統合脳科学研究チーム テクニカルレポート, No 14-01
http://www.itn.brain.riken.jp/05_publication/J_technicalreports/14_BSI_ITN_TechReport_No%2014-01
... 脳を報酬（価値）予測装置とみなした場合の、脳の計算理論概観

4. 講演：全脳アーキテクチャを支えるプラットフォーム（理化学研究所 生命システム研究センター チームリーダー 高橋恒一氏）

※ 講演資料は <http://www.sig-agi.org/wba/10> 上に公開

冒頭に、近々発足予定の全脳アーキテクチャを支援する NPO 法人「全脳アーキテクチャ・イニシアチブ」の設立趣旨・運営体制などが紹介され、引き続き全脳アーキテクチャ実装のプラットフォームとして開発が進められている脳型計算基盤ソフトウェア BriCA の基本コンセプト・機能・開発状況が説明された。

《NPO 法人全脳アーキテクチャ・イニシアチブの意義》

全脳アーキテクチャ実現までの期間を考えると、BESOM、複合学習、Deep Learning の研究は大分進展しているものの、全脳アーキテクチャのレベルにはまだ達していない。大体 10 年から 15 年ぐらいかかるものと見ている。すると、これは人が育つ時間スケールであって、通常の研究テーマより長い時間スケールで研究活動を支える、公益的組織が必要となる。それが「全脳アーキテクチャ・イニシアチブ」である。

《全脳アーキテクチャ中心仮説を構成する試み》

全脳アーキテクチャの中心仮説

脳はそれぞれよく定義された機能を持つ機械学習器が一定のやり方で組み合わせられる事で機能を実現しており、 それを真似て人工的に構成された機械学習器を組み合わせる事で人間並みかそれ以上の能力を持つ汎用の知能機械を構築可能である
--

を実際に造作してみることで構成論的な仮説検証を行う、という構想により BriCA (ブライカ) は開発されている。仮説を以下のように 3 つのサブ仮説にブレイクダウンしてみると見通しがよくなる。

1. 脳のモジュール性

脳神経系ネットワークがランダム結合なのか局所的な結合性が強い（→モジュール性へ）のかは議論のわかれるところだが、全脳アーキテクチャ中心仮説はモジュール性を仮定している

2. 脳器官は機械学習器か

全ニューロンの結合をすべて陽に書き下さないと機能発現しないという立場を取らず、計算論的に一貫性のある単一の原理で動作を定義することで機能を表現可能と考える

3. 機械学習器の非加算性

脳を機械学習器の複合体と捉えたとして、単純な機能の積算で成立しているようには思えない。何らかの非加算性が存在するのではないか？一般化すれば学習が複合した場合に質的創発が起こりうるかという問題だが、ニューロンレベルでの具体的

現象論との対応を図りたい。研究動向としては去年（2014）、Google と Stanford 大のチームが画像を見せるとキャプションを生成するという高次の視覚野に関する研究成果⁶を出している。また、今年の2月には DeepMind 社が教師なしでコンピュータにゲーム攻略をさせる研究を発表⁷したが、Deep Learning と強化学習（Q 学習）を複合させる方法をとっている。

《BriCA の構造と機能》

機械学習を複数組み合わせることが BriCA の基本コンセプトであるが、組み合わせ方は大きく分類して以下の2方式があり、BriCA はいずれにも対応するように設計されている。

1. Composition

互いに独立な機械学習器を2つ以上結合する
→ NeuralTalk@Stanford 大の方法

2. Hybrid

2つ以上の機械学習アルゴリズムを組み合わせる新たな機械学習アルゴリズムを構成する
→ Deep Q-network@DeepMind 社の方法

結果的に、BriCA はハードウェア層から UI 層にいたる6層構造になっている。

《BriCA のモジュール間通信様式》

BriCA を構成するモジュールは I/O を持っており、お互いに通信をする。各モジュールは並列に計算（発火）してその結果をやり取りすることになるが、通常の電子計算機のように共通時刻で同期を取るわけではないので工夫が必要になる。

一つはモジュールの I/O を非同期にコミュニケーションさせる仕組みであり、もう一つはモジュールが一まとまりの計算を完結させて初めて出力側ポートの値を更新する仕組みである。

《BriCA のデータの持ち方》

⁶ レポーター註：

Google のニュースリリース

<http://googleresearch.blogspot.co.uk/2014/11/a-picture-is-worth-thousand-coherent.html>

Stanford のニュースリリース

<http://cs.stanford.edu/people/karpathy/deepimagesent/>

⁷ レポーター註：

Google のニュースリリース

<http://googleresearch.blogspot.jp/2015/02/from-pixels-to-actions-human-level.html>

単純な“モジュール”だけで構成するならば、モジュール間の情報のやり取りはニューロンの発火頻度に対応した数値ベクトルになる（ボトムアップによる究極的な完成形では数値ベクトルで事足りるはずだ）。しかし、認知アーキテクチャのような高次機能も指導原理の一つとして実装する場合は、例えば言語表現のパーズングに必要なグラフのようなデータ構造も取り扱うことが求められる。後者の様式にも対応すべく「Component」と称する部品も合わせて用意している。

Component は非ベクトルポートを持ち、RPC による異なるアドレス空間へのアクセスが可能。

《BriCA の性能》

通信性能という点では、ニューロンの発火頻度上限を 1kHz とおいたときに無視できるレイテンシレベル ($\ll 1/1\text{kHz} = 1\text{ms}$) として 10us 程度がターゲットになる。既にプロトタイプでは 100ns を実現。

一方、人間の脳における処理速度 1PFLOPS を実現するために、コアあたり 10GFLOPS として 100k コアの集積を目指している。現状の H/W としては厳しいレベルで、GPGPU・ニューロチップなど、実行機構のソフトウェア面での進展も見据えた設計が今後の課題になる。

《今後の予定》

いわゆる“バージョン3の法則”（プロト → 汎用化（肥大化） → 最適化）にのっとり、プロトを 2015 年中に、多言語対応を含む完成形を 2016 年以降に、さらに AGI（汎用人工知能）実装系としてきちんと機能するに至るまで、段階を踏んでいく予定。

主な参考文献（レポーターによる追記）：

- 高橋恒一, 板谷琴音, 中村政義 他 (2015) 「認知コンピューティングのための汎用ソフトウェアプラットフォームの設計と開発」, 人工知能学会全国大会 (第 29 回)
<https://kaigi.org/jsai/webprogram/2015/pdf/2I4-OS-17a-1in.pdf>
... BriCA のコンセプトと開発状況
- Song-Chun Zhu (2013) “Statistical Computing and Inference Ch4 Classic Parsing Algorithms”, UCLA lecture, chapter 4
http://www.stat.ucla.edu/~sczhu/Courses/UCLA/Stat_232B/Handouts/Ch4_chart_parsing.pdf
... Parsing 法の機能とデータ構造 (Chart) について
- Paul A. Merolla et al. (2014) “A million spiking-neuron integrated circuit with a scalable communication network and interface”, Science, Vol.345 no. 6197 pp. 668-673
... IBM によるニューロチップ (TrueNorth) 開発報告

5. 講演：人工知能・ロボット次世代技術開発（経済産業省 研究開発課 岡本洋平様）

まず昨年から今年にかけて政府主導で議論を進めてきた「ロボット革命実現会議⁸」の内容が紹介された。

その中では、IoTによるデータ吸い上げと実社会への作用の間にある“処理”の中核を人工知能が担うと位置づけており、NEDO POSTのプロジェクト「次世代ロボット中核技術開発⁹」の中心的テーマになると想定している。

基本計画として大規模実世界データに対する世界トップレベルの研究を実施するために重要なことは“集中的に”研究開発を行っていくことであり、この全脳アーキテクチャのような枠組みから多くの研究者に関わってほしいとメッセージがあった。

⁸ レポーター註： 同会議の公式サイト参照
<http://www.kantei.go.jp/jp/singi/robot/>

⁹ レポーター註： 同プロジェクトの公募要領参照
http://www.nedo.go.jp/koubo/CD2_100040.html

6. 講演：汎用人工知能に向けた認知アーキテクチャが解決すべき知識の課題（国立情報学研究所 情報学プリンシプル研究系 准教授 市瀬龍太郎氏）

※ 講演資料は <http://www.sig-agi.org/wba/10> 上に公開

汎用的人工知能（AGI）が持つ構造を構成論的に理解したいという目的意識から、人間を現状稼働している最高の AGI と捉え、その認知アーキテクチャを説き起こすという論理で議論が展開された。

汎用人工知能（AGI）は、個別の文脈において個別の知的振る舞いをする特化型 AI（Narrow AI）とは反対の概念であり、様々なスキルを習得しようという意味で「汎用（general）」な AI である。また、人間並みの人工知能や心を持つという意味で「強い AI」という概念とも似て非なる概念である。歴史的には、もともと AI は「汎用」を目指していたが、非常に難しいことがわかり、個別課題に特化し、細分化した Narrow AI として広がった。近年は様々な研究成果から再び汎用 AI への興味関心が高まっており、目的を定めない GPT（汎用目的技術）として可能性が探られている。

《汎用人工知能（AGI）の要件と認知アーキテクチャ》

AGI の要件は、AGI 製作者が想定していなかった状況やタスクに対応可能であり、転移学習と汎化能力を持ち、結果的に人間レベルの汎用的知能を持つことである（ただし人間が最高・最善の汎用知能であるとは限らない）。では人間の知能とは一体何であるかということを考えたいが、そのためには環境から耳や目のようなセンサで知覚し、手足を通して環境に働きかけ、知的な反応をするという“人間のモデル化”をする必要がある。人間のモデルは大きくハードとしての「身体」と、ソフトとしての「心理」に分けられるが、後者を支える人間の知的動作には何かの内部構造が必要であり、それを「認知アーキテクチャ」と呼んでいる。

認知アーキテクチャには様々な要素（記憶、学習、問題解決、認識方法、実装...etc.）があって、脳の計算様式も含めて取り入れるのであれば、全脳アーキテクチャのような枠組みと融合させる必要がある。それを「全脳認知アーキテクチャ」として、概念実証から進めている。

《“記号的知識”と“分散表象”の問題》

一般的な認知アーキテクチャと全脳アーキテクチャを融合させる場合、知識記述を変換する必要がある。認知アーキテクチャを考えるだけであれば人間でも理解可能な“記号的知識”を前提することになるが、全脳アーキテクチャは脳型 AI である以上、ニューロンの結合強度などに知識が埋め込まれる“分散表象”型で知識が表現されなければならない。従って何らかの変換機構が必要となってくるが、そこではピアジェの知識発達に関する理論が参考になる。

ピアジェによると、人間は概念自体の形成・概念間の基本関係を学習する「前操作期」

を経て、関係に対して操作を加える方法を学習する「具体的操作期」に至るとされている。その2つのプロセスを分散表象の形から実装することを試みている。脳型AIの知識構造としてはニューラルネットワーク言語モデルが用いられ、単語の出現パターンの学習が成功している。その例が、Mikolovらによる“Skip-gram モデル¹⁰”である。Skip-gram モデルをベースに知識構造を逐次的に構築し、関係する概念を構造化することを狙って市瀬らにより“構造化知識学習”が行われている。

《構造化知識学習》

単語ベクトルの次元数を 300 次元、コーパスとして Wikipedia（英語版）を利用、単語間関係性の初期値として WordNet の is-a 関係を与える設定で実験を行った。最近傍のみを採用する評価基準だと（問題設定上）未知の単語間関係性の同定が困難であったが、上位 100 位以内にまで緩めると 4 分の 1 程度の単語について is-a 関係を正しく同定できており、この手法の可能性が確認されている。

計算上の記号処理と脳の中の事象をどのように結ぶのか、今回のニューラルネットワーク言語によるアプローチで手掛かりが見えてきていると考えている。

¹⁰ レポーター註： 原著論文参照

Mikolov T, Sutskever I et al. “Distributed Representations of Words and Phrases and their Compositionality” Advances in Neural Information Processing Systems 26 (NIPS 2013)

<http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>

7. 講演：感情モデルと対人サービス（玉川大学 脳科学研究所 教授 大森隆司氏）

※ 講演資料は <http://www.sig-agi.org/wba/10> 上に公開

《「相手が何を思っているか推定する」こととは？》

AI を対人サービスに適用し、それが人間にとって満足いくものとなるためには、(AI が) 人間の“価値”を正確に評価 (= 計算) できることが求められる。人間における価値は感情の形で顕現することがあり、その意味で感情の理解や、感情推定手法の開発が必要になる。本講演では子どもの遊びを対象にした観察実験を通して、感情の推定とそれを受けた介入行動という形でのコミュニケーションにどのような動的構造があるのか論じられた。そして、ロボット (AI) が「サービス」するためには「価値の計算」が必要である可能性を示した。

《情動・感情の機能》

感情は進化の過程で発達してきており、全ての動物に感情らしきものが備わっていると言われている。感情には何らかの重要な意義があると考えられるが、その本質的機能は認知したモノ・コトに対する価値の評価にある。戸田 1992 では「感情は、多様な場面での価値計算」と論じられている（参考文献：感情 - 人を動かしている適応プログラム（認知科学選書 24、共著：高田洋一郎、1992 年））。例えば、「嫉妬」は、他者の成果に対する攻撃的姿勢のことだが、自己の価値を相対的に高めるための行動とすれば説明がつく。このような現象論的解釈から「情動・感情＝価値システム」仮説が考えられた。情動・感情は怒り・幸福・悲しみなどいくつかの基本構成要素から成り立っているが、それぞれが評価結果に対応し、脳を強化学習器とみなした時の報酬系として機能している。

《幼児 vs. ロボット 幼児 vs. 保母さん》

これまでの研究では幼児と自律型ロボットのインタラクションを測定するということでは、幼児の反応タイプとして「珍しいもの」か「付き合うもの」へ反応と 2 種類があることがわかってきた。どうしてタイプが分かれるのかと研究を続行した。次に、自律型ロボットが幼児を相手に 30 分間トランプゲームをする実験系を組んだ。最初ロボットに興味があった幼児（全体の 7 割程度、残り 3 割は遊び不成立）も最終的には飽きてしまう。一方で幼児と保母さんが同じトランプゲームをする場合は、飽きさせることがなく、100%の幼児が楽しく遊び終えて帰る。そこで、ロボットにも簡単な構造モデリングをして“つまらない”状態に対して遊びの切り替えを図ると、飽きずに遊ばせることができた。

人間とはすごいもので、熟練した保母さんは最初興味がない子どももほぼ 100%遊びへ引き込むことができる。その能力の源泉は結局のところ、人（子ども）の感情・緊張度・性格などからなる包括的心的状態を推定し、それに見合った遊びを提供する「戦

略的インタラクション」にあるということがわかってきた。そこで、幼児が保母さんが操縦するロボットとゲームをして、プレゼントをもらって終わるという実験を行った。保母さんが操縦するロボットは 100%幼児とのインタラクションに成功した。行動・対話・心拍、子供の動き、全体画像を計測し、分析したところ、保育士が子供の緊張度合い、性格傾向を推定しながら、遊びの状態遷移をさせ行動選択していることがわかった。「相手の心の状態をみて誘導している」これが「遊び」であることがわかった一方、このことを可能としているものは何だ？という疑問を得た。そこから、インタラクションとは心的ダイナミクスであり、心の内部状態は「知的」状態ではなく、「感情」ではないかということである。

《コミュニケーションが成立するベースとしての感情》

心的状態には色々な要素があるが、コミュニケーションの文脈の中で決定的なのが「感情」であることを見出した。感情が相手を受け入れる状態になっていなければ遊び自体が成立しない。これは大人の世界に引きつけても納得できる結果である。「情動・感情＝価値システム」が解明されなければ、ロボットは「サービス」できないということがわかってきた。

今後は他人の感情推定、さらには自分の感情表出を人工知能に組み込んでいくことが、ロボットによる対人サービスの実現には鍵となる。

《質疑応答》

Q: 大人の世界にロボットを導入すると、どうしてもロボットを見下すような傾向があり、対等なコミュニケーションにならない。結果、目的の機能が果たせないことがある。何か解決方法はあるだろうか

A: 大人は難しい。楽しくなくても「楽しかった」と演じることができるので、評価と改善のサイクルを効果的に回していけないことに原因の一端がある。大人、さらには（今後社会的要請が高まると予想される）高齢者向けの対人ロボットサービス開発はこれからの課題と認識している。

主な参考文献（レポーターによる追記）：

- 岡ノ谷一夫（2013）「情動研究の方向性」遺伝, 67 (6), 697-700
... 今後の情動研究の方向性について議論
- アントニオ・R・ダマシオ著 田中三彦訳（2005）『感じる脳 情動と感情の脳科学 よみがえるスピノザ』ダイヤモンド社
... 価値を迅速・的確に計算するシステムとしての脳、という枠組みの紹介
- 浅田稔（2014）「情動発達ロボティクスによる人工共感設計に向けて」日本ロボット学会誌 32 巻, 8 号

<http://www.er.ams.eng.osaka-u.ac.jp/Paper/2014/Asada14b.pdf>

... 情動の人工的な構成方法を提案

- 大森隆司（2015）「汎用人工知能が感情を持つことは有用か」，人工知能学会全国大会（第 29 回）

<https://kaigi.org/jsai/webprogram/2015/pdf/2I5-OS-17b-2.pdf>

... 本講演で紹介された、ロボットと子どもの遊び実験結果が報告されている原著論文

8. 報告：若手の会の活動報告（法政大学 島田大樹氏）

※ 報告資料は <http://www.sig-agi.org/wba/10> 上に公開

若手の会の概要紹介・これまでの活動（各種勉強会）・今後の活動予定が報告された。自分が若手に該当するかどうか？という質問をよく受けるが、特に年齢制限があるわけではなく、若手のやり方についてこられる方や若手を応援していただける方に参加してほしい、とのこと。

会の目的は、基礎知識の共有と若手研究者への情報発信である。2014年現在、フェイスブックで490人程度の参加者がいる。運営は15人程度で、関東と関西それぞれで独自に活動をしている。関東では、Deep Learning と Machine Learning の勉強会を実施し、関西では、計算論的神経科学勉強会を実施している。今後は6～7月では、関東ではベイジアンネット、視覚野について幅広いトピックから汎用人工知能を考えるための勉強会を行う。関西では、この5月にNumpyによる計算論的神経科学入門という勉強会を開催する。また、9月に「全脳アーキテクチャ キックオフハッカソン」を慶應大学日吉キャンパスで開催予定である。

これまでは勉強会による知識共有、ネットワーキングを主体に活動してきたが、今後それは継続しつつ、さらに全脳アーキテクチャ研究自体への貢献も目指している。

《会場からのコメント》

大脳生理学、認知心理学の分野も積極的に勉強してもらいたい。医師と会話ができるということもあるが、一番の効用は「制約がある」こと。何らかのモデルを造るとき、むしろ制約があった方がアイディエーションし易い。若手の会の今後に期待している。

主な参考文献（レポーターによる追記）：

- 全脳アーキテクチャ 若手の会 公式 web サイト
<http://wbawakate.jp/>
... イベント情報など
- 第一回 WBAI ハッカソン
<http://wbawakate.jp/events/hackathon2015/>
... ハッカソンの募集要項やスケジュールなど
- 「全脳アーキテクチャ勉強会若手の会」参加者募集ページ
http://www.sig-agi.org/wba/wakateno_kai_announce-1
... 若手の会に参加したい方向けのご案内

9. レポーター後記

全脳アーキテクチャはある意味シンプルな指導原理（仮説）の下、脳を模倣しつつさらにそれを超える知能機械を志向するものであるが、今回5名の先生方による全脳アーキテクチャの現状報告とも言える講演を聴いて、まだ道のりは険しいものの着実に前進している印象を受けた。機械学習モジュールを低レイテンシで自律分散的に連携させるプラットフォームである高橋先生の「BriCA」、汎用的機械学習モジュールとしての大脳皮質の構成論的実装を追求する一杉先生の「BESOM」、あるいは松尾先生の **Deep Learning** に関する研究、いずれもパフォーマンスの向上を図りながら人間の脳における知的処理に近づいているように思える。また、市瀬先生の知識表象機能研究、大森先生の（価値評価システムとしての）感情機能研究も合わせると、まさに自律的に目的関数最大化を目指して時々刻々意思決定をする主体の姿が浮かぶようである。ここでやはり大きなチャレンジになるのは、フレーム問題・シンボルグラウンディング問題に代表されるような、世界のメタ構造を如何にして自分自身の知識として獲得し、思考・行動の駆動に組み込んでいくかということだと感じた。素過程を拡大しても所詮“fire together, wire together”の Hebb 則によるパターン認識システムが観察されるに過ぎないところ、「その仕方はよくわからないが」パターンを認識するための「最適なお膳立て」をそのパターン認識システム自身が担わなければ脳の機能に迫ることはおぼつかない。

松尾先生によれば **Deep Learning** もいよいよそのブラックボックスで起きている現象自体の理解にメスが入り始めたとのことだが、今回講演された先生方のような様々な階層からの多面的アプローチによって、知能という得体のしれない機能についての理解が深まっていくことを期待したい。

以上