

dwango

DWAL-TR-2016-001

Mar. 2 2016

第12回
全脳アーキテクチャ勉強会
「脳の学習アーキテクチャ」
報告書

生島 高裕*1 佐藤 洋平*2 吉岡 英幸*3 山川 宏*4
Takahiro Ikushima Yohei Sato Hideyuki Yoshioka Hiroshi Yamakawa

*1 株式会社 数理先端技術研究所
Mathematical Science Advanced Technology Laboratory Co., Ltd.

*2 オフィスワンダリングマインド
Wandering Mind

*3 株式会社ナレッジサイン
Knowledge Signs.Co.,Ltd.

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

本報告書の概要

全脳アーキテクチャの構築においては、脳における多様な学習能力に対する理解にもとづいて、機械学習を結合して認知アーキテクチャとして実装する必要がある。そこで今回の勉強会では計算論的神経科学の第一人者である銅谷賢治先生をお招きし「脳の学習アーキテクチャ」というタイトルでご講演いただいた。引き続き、全脳アーキテクチャの関係者とともに神経科学を全脳アーキテクチャの研究開発にどう活かしてゆくかなどについてパネル討論を行った。主に計算論的観点と記号的観点について議論が行われた。最初に、分散表象の記号的な解明、言語の統語論と意味論について精緻な数理的モデルの知見を踏まえた、言語野の機能再現提案があった。次に、タスク、モジュール、ファンクションについての、用語の定義を考えながらの討論、モジュールの生成と検証、確率モデルを生成する確率モデルなどの紹介を交えた議論がなされた。

1. ご挨拶 (トヨタ自動車株式会社 山下勝司氏)

2. オープニング (ドワンゴ人工知能研究所 山川 宏氏)

【要旨】

全脳アーキテクチャ勉強会は、2013年12月開催の第1回から数えて第12回目にあたる。そして、2015年8月にNPO法人全脳アーキテクチャ・イニシアティブ(WBAI)が創設されてから2回目となる。

山川氏のオープニングトークは、これまでの全脳アーキテクチャ勉強会、および全脳アーキテクチャ・イニシアティブとしての活動をふり返り、今後の研究の方向性を示唆するものとなった。

なお、固有の団体やイベント名、用語については、解説と関連情報を参照できるURLを注釈に付記した。

【全脳アーキテクチャ・イニシアティブとしての活動の内容】

全脳アーキテクチャ・アプローチは、脳全体のアーキテクチャに学び、人間のような汎用的な人工知能を創ることを目的とした研究アプローチであり、その基本的な考え方は、脳の各器官を機械学習モジュールとして開発し、それらのモジュールを統合した認知アーキテクチャを構築していく、というものである。特にプラットフォームを共有しながらオープンなコミュニティで構築していくことをめざしている(OpenWBA)。

各モジュールをアーキテクチャ上でマクロな形で統合していくためには、各モジュールを開発していくエンジニアの開発コミュニティが必要であり、さらに、機械学習の専門家や、脳全体をマクロに統合する認知アーキテクチャ研究の専門家、神経科学の専門家など、さまざまな分野の専門家との連携が必要となる。

NPO法人としての全脳アーキテクチャ・イニシアティブ(以下、WBAI)は、その活動を推進する母体であり、汎用的な人工知能の開発ターゲット時期を2030年と置いている。

2015年のWBAIの主な活動としては、4月にNPO法人設立の申請を行い、5月には、各コミュニティによる分散開発を支える、BriCA¹⁾等のプラットフォームを軸としたオープンな開発戦略を明確にした。

また、9月にはWBAIと全脳アーキテクチャ若手の会の共催で5日間の合宿形式でハッカソン²⁾を開催した。課題を「複合学習器の開発」とし、大学生・大学院生を中心とした7チームが参加し、学生や研究所との新たなネットワークづくりの場となった。

11月にはBICA(Biologically Inspired Cognitive Architectures)というコミュニティが主催するカンファレンス BICA 2015(11月6~8日リヨンにて開催)³⁾に参加し、これまでのWBAIの活動や成果を報告するワークショップのプレゼンテーションを行った。

さらに12月には、PayPalやTesla Motorsなどの創業者として知られるElon Musk氏が中心になって米国で設立した非営利の研究機関「OpenAI」⁴⁾に対して、WBAIの理念との親和性が高いとして、歓迎のコメント⁵⁾を発表した。

汎用人工知能の研究を、実装の対象とする脳機能の範囲と、研究アプローチの2軸のマトリクスでマッピングすると、WBAIは、価値システムを含む脳全体の機能の実現を支援しており、より脳に忠実な実装のアプローチを試みるグループに位置する。これまではこのグループのプレイヤーは比較的少なかったのだが、ここ1年ほどは、Deep Mind等の参加者が増えてきている。

このような環境の中、WBAIは、賛助企業などの支援を受けながら、研究者のネットワークによる汎用人工知能の研究開発を支援/推進するために、積極的な情報発信や、学習環境の構築、汎用人工知能の評価指標や統合プラットフォームの提供などをしている。また、脳をモデルにした汎用人工知能の開発では、神経科学との融合が重要となるため、本勉強会を含め、神経科学の最新の知見との接点を持つことを重要視している。

【今後の全脳アーキテクチャの研究の方向性】

直近の人工知能を取り巻く技術的進展を見ると、NIPS(Neural Information Processing Systems)⁶⁾の2015年のカンファレンスの内容からも、深層学習の適用範囲が広がっており、人工ニューラルネットワークで実現可能な領域が広がってきたことが窺える。

神経科学分野においても、コネクトームから認知アーキテクチャにヒントを与える萌芽的研究が進んできていることが実感できる。

前述のように、全脳アーキテクチャのアプローチは、機械学習器の開発と、それらをマクロに統合する認知アーキテクチャの構築を組み合わせたものであるが、機械学習の分野で、大脳新皮質をモデルとした深層学習の研究が進んできたことと、神経科学の分野で、コネクトームなど、脳全体の結合様式の研究が進んできたことは、このアプローチの大きな追い風になっている。

認知アーキテクチャは、脳が各器官をマクロなネットワークで結ぶことで複雑な認識行動を実現するのと同様に、各機械学習器をコンポーネントとみなして、その配置を描いた設計図のようなものである。

神経科学の分野で研究が進んできたコネクトームは、脳全体の結合様式と言えるもので、コネクトームをモデルに認知アーキテクチャを構築することで、全脳アーキテクチャの実現に道筋が見えてきた。

さて全脳アーキテクチャは大雑把に言えば、生物が進化で獲得した部分をアーキテクチャとして設計し、発達や学習で獲得する部分を機械学習で実現するアプローチといえる。

さて本日のテーマは、「脳の学習アーキテクチャ」について考えてみると、これは学習とアーキテクチャということばが一体化している。そこで本日のこの後の講演やパネル討論では、全脳アーキテクチャの構築に向けて、どの部分を学習し、どの部分を設計すべきか、という問いに深く関わっていけるのではないかと期待している。

【質疑応答】

参加者からは、「認識」における脳の活動だけではなく、「判断」、「行動」における脳の活動についても人工知能の観点でメカニズムが解明されているのか、という質問があった。これはまさに、今回のゲスト講演者である沖縄科学技術大学院大学 銅谷賢治教授の強化学習に関する講演のテーマと合致しているため、銅谷氏の講演の絶妙な導入となった。

【主な参考文献】

1) BriCA は、ドワンゴ人工知能研究所と理化学研究所が中心になって開発が進められている、全脳アーキテクチャにおける分散開発を支える汎用ソフトウェアプラットフォームで、任意の数の機械学習器モジュールを階層的に結合した認知アーキテクチャを構成可能としている。BriCA については、2015 年 7 月開催の第 10 回勉強会で詳細が紹介されており、WBAI の HP でも資料が公開されている

□第 10 回勉強会における BriCA に関する発表資料

<https://docs.google.com/a/brainvalley.jp/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWVpbmNzaWdhZ2kxNHxneDo3YmJjOGZlYWE1YTTFhOGRj>

□WBAI の HP で公開されている BriCA に関する資料

<https://kaigi.org/jsai/webprogram/2015/pdf/2I4-OS-17a-1in.pdf>

□BICA2015 で発表された BriCA に関する資料(英語版)

http://wba-initiative.org/wp-content/uploads/2015/05/BICA_2015_submission_9.pdf

2) □WBAI 主催ハッカソンに関する詳細は、以下のレポートを参照

<http://wbawakate.jp/posts/events/%E7%AC%AC1%E5%9B%9Ewbai%E3%83%8F%E3%83%83%E3%82%AB%E3%82%BD%E3%83%B3%E6%B4%BB%E5%8B%95%E5%A0%B1%E5%91%8A/>

3) □BICA2015 の模様

<http://bicasociety.org/meetings/2015/>

□BICA2015 における各プレゼンターのプレゼンテーションビデオ

<http://bicasociety.org/videos/bica2015.html>

4) □Open AI の Web サイト

<https://openai.com/blog/introducing-openai/>

5) □WBAI の Web サイトで発表されたコメントの内容

<http://wba-initiative.org/en/923/>

6) NIPS(Neural Information Processing Systems)は、神経情報システムに関する先進的な研究に従事する非営利の研究団体。カナダに本部を置き、2015 年のカンファレンスは、モントリオールで開催された。

□NIPS(Neural Information Processing Systems)の Web サイト

<https://nips.cc/>

3. 脳の学習アーキテクチャ (沖縄科学技術大学院大学 銅谷賢治氏)

【要旨】

現在人工知能の開発が急速に進められている。人工知能開発と脳科学研究は相互に影響を与えながら共進化的に発達してきた背景がある。今回の講演では脳と人工知能をテーマに最新の研究の動向を交えつつ解説を行う。脳と人工知能の違いというのは現在では相対的なものになりつつあり、脳特有の機能と考えられていた自主性、情報選択能、能動性、モジュール自己組織化というものも人工知能において実装可能なものになりつつある。

【はじめに】

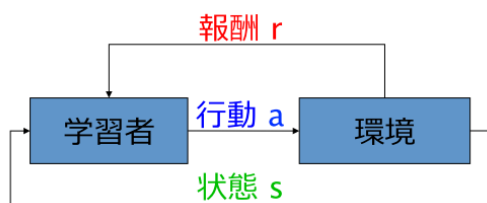
人工知能開発の方向性については多くの議論がなされている。一つの立場は脳そのものに学ばべきというものであり、もう一つは機能を実現することが大事であり、必ずしも脳そのものに近づく必要はないというものである。過去を振り返ると、この二つの研究分野は共進化的に発達してきたことがわかる。研究開発の流れとしては脳を模した学習の強みが認識されてきており、脳そのものに学んだほうが良いのではないかと考えられてきている。

しかしながら脳と人工知能の違いとはどのようなものなのだろうか。これらの違いは大きくは四点に集約されると思われる。ひとつは自主性である。脳というのは自主的に出来上がってきて様々なことができるようになるが、人工知能の多くは何らかの形でプログラマの手を経る必要がある。もうひとつは情報選択という点である。脳は状況に応じて入力や出力、目標の選択を自律的に行うが、現在の人工知能の多くはプログラマによって何を情報選択するかという点については事前に作りこまれる必要がある。さらには能動性という違いもある。人間の脳はトップダウン的に仮説を検証するような形で、少ないサンプルデータから特徴を把握することができるが、人工知能はボトムアップ的に大量のデータを処理するという過程が必要になる。またモジュールの自己組織化という点での違いがある。人間の脳は状況に応じて過去に獲得した様々なモジュールをうまく組み合わせることで運用することができるが、現在の人工知能では状況に応じてモジュールを自律的に組み合わせるということは難しい。

このように脳と人工知能は様々な違いがあるが、これらの違いが現在どのようなものになっているか説明を行っていきたい。

【人工知能研究】

人工知能研究で重要になるのが強化学習という概念である。強化学習モデルでは、学習者が報酬の予測を行い、その報酬予測を基に行動を選択し、行動結果が学習者にフィードバックされて予測と結果の誤差から学習者が学習していくことが考えられている。



この強化学習モデルと他の手法を組み合わせることで人工知能が自律的に問題解決方法を発展させていることが多くの研究から示されている。

一つの例として起き上がり運動を学習するロボットがある(□1)。この研究では、起き上がり時の頭の高さを報酬に、転倒を罰としてプログラムを行い、数百回ロボットに学習をさせた結果、ロボットは自ら起き上がるようになった。この研究で難しかったのが報酬や罰のパラメータ設定である。例え

ば罰を大きくし過ぎると転ぶのを恐れ、じっと倒れたままになるという局所最適解に陥ることもある。また現在の報酬と将来の報酬のどちらに重きをおくかという変数が重要になるが、これも事前に調整する必要がある。この研究では起き上がり運動のような非線形課題にも強化学習が適用できることを示すことができた。しかし報酬のデザインや学習の仕方の学習については人の手を経る必要があるという問題が残された。

このような問題意識をもとにサイバーローデントの研究が進められた(□2)。この研究では動物の進化発達を模したものであるが、一群のロボットに生存と複製を報酬として行動させ、その中で報酬関数がどのように変化するかについて検証を行った。各ロボットには基礎的な行動や報酬、さらにはどれくらい先の報酬を予測するかというパラメータをベクトルとして、生存行動(バッテリーパックの獲得)や複製行動(他のロボットとの接触によるパラメータのコピー)を繰り返し行わせた。これに寿命、突然変異、セレクションという要素を加えて世代による変化を調べたのだが、世代を経ることで他のロボットを効率よく認識できるパラメータが生成されることが示された。強化学習だけであれば事前に報酬関数やパラメータチューニングの調整をする必要があるが、強化学習と進化計算を組み合わせでこれらの作業をロボット自身が自律的に行うことが示された。

しかしながらこのサイバーローデントの研究においても状態表現については人の手を経る必要があった。この状態表現とは対象物そのものをどのように表現するかというものである。この研究の場合、ロボットに対象物がどれくらいの距離や角度にあるということを入力せよということ事前に作りこむ必要があった。しかし近年 Google deep mind のグループが発表した Deep Q network では、ゲーム画面そのものを入力情報とし、その中の必要な情報を4層ネットワークで特徴抽出して学習させ、人間以上の実力を得るに至った(□3)。つまりディープネットワークと強化学習の組み合わせで、状態表現そのものを自律的に獲得できる可能性が示された。

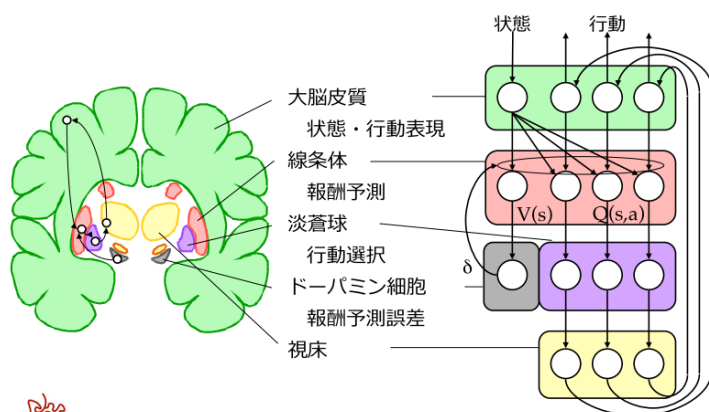
これらの研究から示されるように現在では強化学習に様々な手法を組み合わせることで人工知能に自律的な学習機能を組み込むことができるようになってきている。それではこの強化学習について脳科学分野ではどの程度まで明らかにされてきているのだろうか。

【脳科学研究】

この分野の代表的な研究として猿の脳のドーパミン神経細胞と報酬予測誤差の関係について調べたものがある(□4)。強化学習理論で用いられる重要な指標の一つに TD 誤差というものがある。これは予測した報酬と実際の報酬、今後さらに予測される報酬の三者間の誤差を現す信号であるが、この研究で示されたドーパミン神経細胞の反応は TD 誤差と見事に一致していた。この研究結果からドーパミン系が強化学習に重要な役割を果たしていることが考えられるようになった。ではこのドーパミン系は脳の中ではどこに最も投射されているのだろうか。

脳の中で最もドーパミン投射を受けるのは線条体と呼ばれる領域である。解剖学的にはこの線条体は基底核の一部を構成し、視床や淡蒼球と接続して大脳皮質とループ構造を形成していることが知られている。

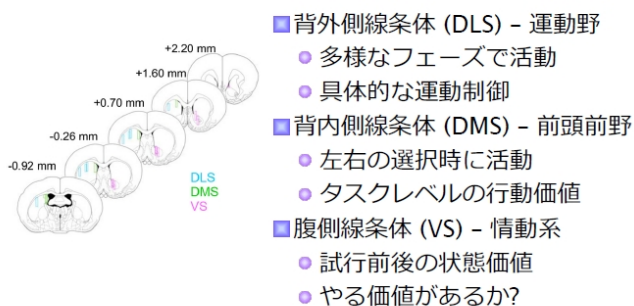
大脳基底核の強化学習モデル (Doya 2007)



この線条体は大脳皮質からシナプス入力を受けているが、このシナプス入力がドーパミン依存的に強化されたり減弱したりすることも知られている。このようなことからドーパミンが大脳皮質と基底核をつなぐ強化学習、とりわけ報酬の予測計算に関わっているのではないかと考えられてきた。

私達のグループではこの仮説を検証するためにラットにギャンブル課題を行わせ線条体の神経活動の変化を調べた(□5)。結果、線条体の背外側領域と背内側領域に行動価値が反映され、腹側領域には状態価値が反映されることが示された。解剖学的には線条体の背外側領域は運動の実行に関わる運動野と接続しており、背内側領域は行動価値計算に関わる前頭前野に接続し、腹側領域は報酬情報を示す情動系と接続している。このようなことから線条体は大脳皮質と接続して階層的で並列的な形で強化学習に関わることが考えられた。

皮質－線条体回路の階層構造

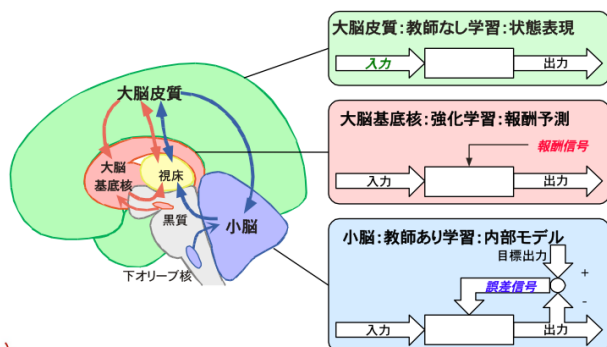


このように線条体は強化学習に関わっているようだが、脳全体のレベルでは強化学習はどのように行われているのであろうか。

脳は多くの部分が階層的につながることによって構成されているが、その中でも大脳皮質と小脳、大脳基底核が学習回路を形成していることが考えられている。

学習アルゴリズムによる分化

(Doya, 1999)



具体的には大脳基底核はドーパミンニューロンを介した強化学習に関わっており、小脳は体性感覚からの誤差信号をもとに教師あり学習と内部モデルの構築に関わり、大脳皮質は多くの入力信号をもとに教師なし学習と状態表現の獲得に関わることが考えられている。このように大脳皮質、大脳基底核、小脳はそれぞれ得意とする学習方法があり分業的に学習を進めているようだが、これらのモジュールが状況に応じてどのように組み合わせられるのかという点については十分明らかにされていない。

このモジュールの組み合わせに関連する話題としてモデルフリーとモデルベースの意思決定というものがある。モデルフリーの意思決定では大脳基底核に蓄えられた価値情報を元に価値を最大化するよう行動が選択される。それに対してモデルベースの意思決定では小脳に構築された内部モデルを用いて脳内シミュレーションを行い、その結果を元に最適な行動が選択される。この二つの方法はそれぞれ一長一短であり、モデルフリーは簡単に行動選択ができるが多くの反復経験が必要になり、モデルベースは柔軟な適応が可能になるが処理が複雑になってしまうという特徴がある。

モデルベースの意思決定では大脳皮質と大脳基底核、小脳を繰り返しつなぐ複雑な処理が必要であるが、一度脳が学習してしまえば必ずしも毎回複雑な処理を行うことなくリアルタイムで効率よく処理できるのではないかとということが理論上予測されていた。それを実証したのが以下の研究である(□6)。

この実験では盤面で駒を動かす課題を行っている時の脳活動を調べているのだが、実際の行動に先立って大脳皮質や小脳、基底核といった脳内シミュレーションに関わる領域の活動が高まることが示されている。

また脳内シミュレーションをニューロンレベルで示した研究もある(□7)。**Redish** のグループは、ネズミが経路の選択をしている時の海馬の様子をデコーディングし、脳内シミュレーションに対応してする海馬のニューロン活動を示した。また現在我々がやっている二光子顕微鏡を用いた研究ではマウスの頭頂葉で脳内シミュレーションにあたる活動が起こることをニューロンレベルで示している。このように、今までは仮説レベルで考えられてきた脳内シミュレーションを現在では実験レベルで掘り下げることができるようになってきている。

また脳のモジュール自己組織化という点についても研究が進んでいる。去年発表された **Koechlin** のグループは、変動する課題へ適応する過程で、新たなモジュールの生成、検証、企画に対応する脳活動が起こることを機能的 **MRI** を用いた研究で示している(□8)。

さらに脳特有の認識機能を人工知能に実装する研究も進んでいる。ヒトの脳は少量のサンプルデータを元にコンセプトを獲得することができる。それに対し従来の人工知能はディープラーニングのようにコンセプト獲得のための大量のサンプルを必要としていた。**Tenenbaum** のグループの研究

では、階層的かつ確率論的な手法でヒト同様に少ないサンプルからコンセプトを獲得しそれを再現できることが示されている(□9)。

以上に示したように現在の脳科学研究では、脳内シミュレーションやモジュール自己組織化といった意思決定に関わるメカニズムの研究が進められており、またそれを人工知能に適用しようとする研究もなされている。

【脳のビッグデータと大規模計算のもたらすもの】

世界的に脳のビッグデータを対象にしたプロジェクトが動いているが、日本でも「革新脳プロジェクト(2014—2023)」と呼ばれるプロジェクトが進められている。これはマーモセットと呼ばれる小型のサルを対象に、マクロレベルからミクロレベルの神経活動を様々な方法で計測し、それを一つの神経回路モデルとして統合しようとするものである。

もう一つのプロジェクトとしては理研の「京」による世界最大規模の大脳皮質回路シミュレーションがある。これは17.3億個のニューロンで約10兆個のシナプスのネットワークをシミュレーションするものであるが1秒間の再現には40分を要している。脳そのものの性能には遥かに劣るが、それを実際にコンピュータでシミュレーションしうる時代になってきている。他の研究としてはラットの脳活動を大脳基底核と視床、大脳皮質の関係でシミュレーションを行おうというものもある。またポスト「京」コンピュータの課題として思考を実現する神経回路機構の解明と人工知能への応用というものがあげられている。

【まとめ】

脳と人工知能の違いというのは実は絶対的なものではない。自主性という面ではサイバーローデント研究のようにロボットに自主性を持たせようとする研究も進められており、情報選択の自律性という点についてはディープラーニングを用いた方法で対応できるようになっている。能動性については、データ生成ができる双方向性のディープネットワークが注目されており、これをデータ生成モデルとして使えるという形で汎用性が出てきている。モジュール自己組織化に関してはよりフレキシブルに必要なモジュールを生成し構成するものとして **Probabilistic programming** という形で定式化が進みつつある。こういったことから人工知能も今後はより脳に近いものになっていくのではないかと考えている。

【質疑応答】

Q1: モデルベースで順モデルがなぜできるかという点ですが、生物学的知見を踏まえてアルゴリズム的な面から説明して頂きたい。

A1: 身体情報と空間情報の二つがサンプルとして与えられれば原理的には教師ありモデルの方法で内部モデルを獲得することができる。頭頂葉は運動前野と相互連絡しており運動情報と空間情報の統合がなされている。また小脳と大脳基底核は頭頂葉と運動前野と相互連絡している。このような脳内のつながりを考えると、小脳で運動情報と空間情報をもとに内部モデルを獲得されるということは原理的には可能だろうと思われる。

Q1: 情動と報酬の関係はどのように考えられていますか。

A2: 情動をどう捉えるかという点だが、工学的な立場からはそれは学習の仕方の制御に関わるものではないかと考えている。脳内では様々な神経修飾物質が働いている。ドーパミンは報酬予測誤差、

アセチルコリンは学習速度、ノルアドレナリンは探索の搾り、セロトニンは予測の時間スケールに関わっていると考えている。情動というのはこれらの学習に関わる大域変数やパラメータの制御に関わっており、この制御が破綻することでうつ病やアルツハイマー病などの症状が現れると思う。

【主な参考文献】

- (□1) Morimoto, J. and K. Doya (2001). "Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning." *Robotics and Autonomous Systems* **36**(1): 37-51.
- (□2) Doya, K. and E. Uchibe (2005). "The cyber rodent project: Exploration of adaptive mechanisms for self-preservation and self-reproduction." *Adaptive Behavior* **13**(2): 149-160.
- (□3) Mnih, V., K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland and G. Ostrovski (2015). "Human-level control through deep reinforcement learning." *Nature* **518**(7540): 529-533.
- (□4) Schultz, W., P. Dayan and P. R. Montague (1997). "A neural substrate of prediction and reward." *Science* **275**(5306): 1593-1599.
- (□5) Ito, M. and K. Doya (2015). "Distinct Neural Representation in the Dorsolateral, Dorsomedial, and Ventral Parts of the Striatum during Fixed and Free-Choice Tasks." *The Journal of Neuroscience* **35**(8): 3499-3514.
- (□6) Fermin, A., T. Yoshida, M. Ito, J. Yoshimoto and K. Doya (2010). "Evidence for model-based action planning in a sequential finger movement task." *Journal of motor behavior* **42**(6): 371-379.
- (□7) Johnson, A. and A. D. Redish (2007). "Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point." *The Journal of neuroscience* **27**(45): 12176-12189.
- (□8) Donoso, M., A. G. Collins and E. Koechlin (2014). "Foundations of human reasoning in the prefrontal cortex." *Science* **344**(6191): 1481-1486.
- (□9) Lake, B. M., R. Salakhutdinov and J. B. Tenenbaum (2015). "Human-level concept learning through probabilistic program induction." *Science* **350**(6266): 1332-1338.

4. WBAI 創設賛助会員プレゼンテーション (株式会社 Nextremer 古川氏)

5. パネルディスカッション「神経科学と全脳アーキテクチャ」

【要旨】

本パネルでは冒頭に市瀬氏および一杉氏から問題提起を頂き、これらについて主に計算論的観点と記号的観点について議論を行う予定であった。

実際の内容は、そのテーマの下、市瀬氏からは、知能内部構造である認知アーキテクチャの重要性と、認知アーキテクチャの一例として **CogPrime** の紹介、脳の機能モジュール化とその評価、分散表象などで実現されている、記号的な知識でのコミュニケーションの仕組みの解明などの課題が挙げられた。

一杉氏からは、脳の目的関数は報酬期待値最大化であること、数理的表現としてモデルベース強化学習を骨組みとすることの提案、人間のタスク実現には汎用事前知識が必要不可欠であること、最近の理論言語学の進展からの、言語の統語論と意味論について精緻な数理的モデルの知見を踏まえた、ベイジアンネットによる言語野の機能再現の提案があった。

パネルディスカッションでは、会場からの質問の下、タスク(人間の行う仕事)、モジュール(タスクを実現する脳の部位)、ファンクション(モジュールの機能)についての、用語の定義を考えながらの討論が行われた。銅谷氏からは、**Koechlin** グループの研究事例から、必要に応じたモジュールの生成と検証、その空間的構造についての議論、評価については、メタ認知などに関してピアレビューによる異常検知、記号処理については、連続と離散問題、**Tenenbaum** グループの研究例から、確率モデルを生成する確率モデルについての話があった。

最後は、なぜ脳のモジュールが作られて来たかについて議論され、大きく盛り上がった。

パネリスト:

沖縄科学技術大学院大学 銅谷賢治氏,
産業技術総合研究所 人工知能研究センター 一杉裕志氏,
国立情報学研究所 市瀬龍太郎氏

モデレータ:

ドワンゴ人工知能研究所 山川宏氏

【問題提起 市瀬龍太郎氏】

<汎用人工知能(AGI)>

全脳アーキテクチャのミッション・ステートメントは、「脳全体のアーキテクチャに学び人間のような汎用人工知能を創る」ことで、ジェネラルな、一般的なところが従来の人工知能と大きな違いである。

従来型の人工知能は、将棋は将棋の人工知能、チェスはチェスの人工知能というようなかたちで、チェスができるような人工知能は、将棋にはそのままもっていけないという問題があった。

汎用人工知能を考えるときに、人間の知能は基本的には汎用であるので、人間の知能を考える必要がある。

<人間の知能>

人間の知能は環境から情報を得、その環境に対して行動を生成することで知能を発揮している。そう考えると、人間は実際に動くハードウェア(体)とソフトウェア(心理)で動いていると考えられる。ここでは、知能の観点から心理のソフトウェアに非常に興味がある。

心理のソフトウェア部分をどのように作るのかについては、知的な動作実行を、完全にモノリシックに作ることは非常に難しい。従って、内部の構造、認知アーキテクチャが必要と、人工知能研究で議論されてきた。

<単純反射エージェント[Russell09] (□1)>

認知アーキテクチャーとは、単純なものでは、「環境からのセンサー入力、現在の状況判断、行動選択(過去の知識利用)、アクチュエータで行動」といったものである。

単純なものでも、内部に構造があり、その構造で知的な行動が実現できる。

<CogPrime(□2)のアーキテクチャ>

CogPrime では、多数のモジュールから構成されている。環境から情報が入ってくると、認識階層では階層的な機構で情報を知識に変換し、AtomSpace で重み・ラベル付きハイパーグラフで知識を保持し、最終的に認知プロセスで学習や推論を実行する。

<全脳型認知アーキテクチャ>

脳の機能のモジュール化を、どのようにして同定していけばよいか問題となる。

脳の機能モジュール化ができれば、そのモジュールの完成度、改善の余地を示せる指標が必要ではないか。実際に、動作をさせることを考えると、人間は言葉など、記号的な知識でコミュニケーションを行ったり、何らかの思考をする。対して脳の中では分散表象などで実現されている。その部分をどうやってつなぐかが、認知アーキテクチャを考えることにおいて、重要な論点になるのではないかと思う。

【問題提起 一杉裕志氏】

<全体アーキテクチャの骨組みの解明が優先度の高い課題>

まず、全体アーキテクチャの骨組みの解明が重要であるが、WBA 勉強会の脳、機械学習の話を通して大体形は見えてきたと思われる。

大事なことは、全体が巨大で複雑な、機械学習アルゴリズムであって、目的関数は、報酬期待値最大化であることである。

その目的を達成するために、脳全体が協調して動いているという見方をしている。

アルゴリズムの基本は、モデルベース強化学習で、脳らしい振る舞いは、行動する前に考えて、考えた結果、意思決定する。

これを数理的に表現したものが、モデルベース強化学習で、それを骨組みとすることによって、ほかのパーツもその部品として、うまく当てはまる印象がある。

大脳皮質と小脳はシミュレータで、海馬は毎回シミュレーションする無駄を省くために、シミュレーション結果をキャッシュするメモ化(Memoization)の役割でないかと思う。このように脳全体を考えていくことが重要だと思っている。

<機械学習の観点から、私が重要だと考える課題>

機械学習アルゴリズムは、局所解・過適合にはまってしまうので、それを抑える事前知識を明らかにすることが大事、また、記号処理と統計的機械学習の統合、昔からのテーマである言語の意味表現と推論方法を解明する必要があると思う。

<汎化性能を上げるには事前知識が必要>

脳については神経科学からいろいろなヒントが得られると考えている。

事前知識に関しては、機械学習の性能を上げるためには、事前知識をつくりこむことしかない。

ノーフリーランチ定理([□3](#))によれば、どんな問題に対しても、汎化性能を発揮するような機械学習アルゴリズムは作れない。この理由で、性能を上げるためには、学習対象の問題領域の知識を、可能な限り作りこむしかない。

汎用人工知能は汎用だから、問題領域というのではないのではないかという風に、誤解されるかもしれない。

<目指す「脳型汎用人工知能」は、「人間のタスクに特化した人工知能」>

目指すものは、脳を模倣した汎用人工知能で、人間のタスクに特化した人工知能なわけである。

それは実は非常に狭くて、ありうるタスクは非常に広くて、そのほんのごくごく狭い一部の領域が、人間が行うタスクで、これに特化しなければならない。これに関する研究はない。

チェス、将棋、画像認識などかなり狭いタスクの研究はあったし、ものすごく広いタスクへの対象とした機械学習はあった。

「パーセプトロンが万能とか、Q ラーニングの収束性が保証されている」などは、すごく広いところを対象としている。

理論的には収束するけれども、現実には動かないというのは、対処に関する事前知識の作りこみが全くないので、実問題に適用すると全く性能がでないということである。

従って目的は、実問題を解ける汎用人工知能であるので、人間が行うようなタスクに、特化した作りこみが必要、そのためには、脳の知見を参考にして、どういう作りこみをされているかを明らかにする必要がある。

< Generic Priors for Disentangling Factors of Variation (□4) >

Bengio も「Generic Priors」で、わりに、なににでも適用できる汎用性の高い事前知識を 10 個ぐらいに分類して説明している。こういうことをどんどん明らかにすることが必要。

個別にはいろいろなシステムで作られているが、全部ひとつの機械学習アルゴリズムに入れたシステムは存在しないし、作らなければ人間なみの汎化能力は得られないと思う。

中身は、

- Smoothness
- Multiple explanatory factors
- A hierarchical organization of explanatory factors
- Semi-supervised learning
- Shared factors across tasks
- Manifolds
- Natural clustering
- Temporal and spatial coherence
- Sparsity
- Simplicity of Factor Dependencies

わりとあたりまえのことがならんでいる。

< 追加すべき汎用事前知識の例 >

概念の関係に is-a 関係がある、記号処理の人工知能では使われている。

人間は猫で経験した知識を、犬などの他の動物にも汎化出来る、例えば、追いかければ逃げる、怒らせると攻撃される、といった知識は全ての動物に共通、もし、猫と犬という概念を別々に保持するシステムだとすると、猫の知識を犬に汎化させるということとはできない。

動物という全体を包含する概念をもっていなければならない。

人工知能の分野では昔から行われており、ミンスキーのフレーム表現で対応していた。

そして、人間が手で知識を書いていた。

統計的機械学習によって、is-a 階層を自己組織化するようなシステムは、まだないのではと思う。

< フレーム表現の例 >

例としては、獣に関する知識として、足が 4 本で、毛がある、それが猛獣という概念、牛という概念にも引き継がれる。このようなものを、自己組織化するようなものが必要とされる。

< 線条体と中脳の間スパイラル構造[Haber 2003] (□5) >

脳の中で、大脳基底核の中でスパイラル構造があって、階層的な強化学習を脳は行っているらしい。脳は、ディープ Q ネットワークとかは違うような気がする。

各階層で、それぞれが独立にも動作することもできるし、適宜、上が下を抑制したり、下が上に割り込みをかけたたり、つねに全体として情報のやりとりを行って、全体が最適化されるような振る舞いを行っているように思っていて、これを実現するには、階層的なベイジアンネットと強化学習を組み合わせたようなシステムが必要とされている。

< 言語野モデルに関する分野 >

産総研の脳型人工知能研究チームは言語野モデルの構築にも取り組み始めている。

言語は脳内の状態を符号化したものであるため、言語が持つ構造は脳の構造を反映していると考えている。

従って、言語を知ることは脳のアーキテクチャを推定するよいヒントになる。

言語野はヒトにしかなく侵襲的実験ができないために、視覚野や運動野に比べると情報処理の理解はあまり進んでいない。

しかし、言語の性質は古くから言語学者が研究しており、いろいろな「言語現象」と呼ばれるものが知られているらしい。

私が特に注目しているのが最近の理論言語学の進展であり、言語の統語論と意味論について精緻な数理的モデルが構築されつつある。

これらの知見を踏まえて、ベイジアンネットを用いて言語野の機能を再現しようと考えている。

【パネルディスカッション】

■山川氏から市瀬氏へ

最近、知能へのアプローチとして、ディープラーニングや強化学習などの機械学習が利用されることが増えてきていますが、市瀬氏から見てどのように思われますか？

■市瀬氏

古典的 AI の観点からすると、記号から逃げられない宿命がある。人工知能を実現したとき、安全性の観点から内部の動作を理解したい要求があると、人間が理解するためには、記号的部分がある程度ないと、理解できない。記号で扱うとなると、離散的になり、離散的なものになると、組み合わせで示されるというところが、古典的な人工知能の立場であったと思う。

■山川氏から会場へ

ただいま記号の話がありましたが、会場から質問などございますか？

■会場から

仮説生成と仮説検証では、思考の仕方が違っていると思われるのですが、最近の研究では、どの程度わかっているのでしょうか？

■銅谷氏の回答

人間はこれまでのルールではうまくいかないと、新しいルールを学習するといったことを行っている。講演で紹介した [Koechlin グループの研究\(□6\)](#) は、この過程が、必要に応じてモジュールが生成されるような、一種の人工知能プログラム(ノンパラメトリックベイズ)でフィッティングできるということを示している。

そして、新しいモジュールが生成されたタイミングで活動する脳の場所、生成されたモジュールが検証される過程で活動する場所があることを fMRI 実験により示している。そこがその論文の面白いところで、仮説生成と仮説検証の脳機構に迫る研究であると思う。

■会場から引き続き

モジュールはサイズが決まっているのか？話題にしているテーマによってサイズはかわるのか？

■銅谷氏の回答

モジュールというのが空間的に特定できるのかという問題がある。

脳の側頭葉の中には顔や道具、風景など特定の視覚刺激に応答する領域が別個にあって、これらは特定の機能モジュールと思われる。顔領域にはたくさんのニューロンがある、例えば、ある特定の人に反応するニューロンがあるが、それらがモジュール内の特定の場所に全部固まっているかという、必ずしもそうではない。

空間的に境界がきちっと分かれたようなモジュールも、あるレベルではあるかもしれないが、例えば山川ニューロンが、この領域に数十個分散している、そういった空間的にオーバーラップした機能モジュールもあり得るのではと思う。

モジュールを1個増やすという場合に、ぐりぐりとあるスペースを空けて配置する代わりに、既存のニューロンの中で、比較的新しいものに近い特性を持ったものを寄せ集めて、フレキシブルに新しい機能モジュールを作ることができる、こんな仕組みが多分あるのではないかと思うが、検証されたわけではない。

■山川氏

特に新皮質において、物理的なコネクション長において近距離結合の割合が高く、この性質はモジュラリティが構成される大きな一因でしょう。

■銅谷氏の回答

あまり厳密なことは知らないが、大脳皮質の回路の抑制性と興奮性の結合が、どれくらいの広がりにつながっているかといったことで、モジュールの空間構造は多分決まると思う。

■山川氏

うまくいっているかどうかの検証は、ある意味モジュール単位もしくは大きな単位で、自分自身でパフォーマンスを評価する、メタな視点が入ってきて、ドラスティックに組み替えるべきかどうかの判断をしていると思う。

もともと、市瀬氏の初めの質問で、メトリックの話がでたが、これは評価という意味ですよ。

■市瀬氏

各々にモジュール分けができたとして、全体を全体最適化することになる。

全脳アーキテクチャで分散開発になると、それぞれのモジュールが、パフォーマンスを出すかを評価しなければならない。

たとえば、強化学習のモジュールがあって、それに対してして、ちゃんと反応できるようなメトリックができれば、それを組み合わせて全体ができる。その辺の各々のところをどういう具合に考えればよいのかなと思う。

■山川氏

各モジュールが常に動いているわけではないので、動いていないときに評価されても意味がない。

マルチエージェントシステム視点からの知見はあるでしょうか？

■銅谷氏

パフォーマンスの評価には、いろいろなやりかたがあり。例えば小脳では、登上線維は運動の目標と結果などの誤差信号を表現していることが知られている。

小脳の内部モデルはフィードフォワードで確定的な予測モデル、大脳皮質はリカレントな確率的なモデルを実現していると考えられる。また確率モデルではトップダウンの予測とボトムアップの信号のマッチングが取れているかの演算がされているはずで、そういったものを使って、予測が正確かどうか、ローカルに評価することもできるはずである。

メタ認知に関しては、実際に働く脳部位と、メタ認知専用に行っている脳部位があるかどうかという疑問がある。

メタ認知を専門とする脳部位があるというよりは、それぞれ専門性の違う専門家が脳内にいて、世の中うまく行っている場合は、各専門家が同じことを言っているけれども、ある状況では各専門家の予測が違って来るので何かおかしいことに気付く。そういったピアレビューによるメタ認知みたいなものの方が、脳では実現されやすそうな気がする。

■一杉氏

ベイジアンネットであれば、同時確率が0になれば、お互いにシステムが矛盾していることになるので、そういったことで検出できるのではと思う。

■山川氏

記号を含めた表現の問題、市瀬氏からは「記号処理と分散表現」、一杉さんからは「記号処理と統計的機械学習」の話があった、銅谷氏は記号処理についてどう思っているか。

■銅谷氏

記号処理にあたって、変数が連続か離散かという問題と、変数を代入できるかという問題がある。連続か離散かだと、連続の確率分布を近似するのに、混合ガウスモデルを使うことが普通だから、これは一種、連続的なものを離散化したと見ることで、比較的簡単に実現できそうに考えられる。

ただ、そうして捉えた、ある種のクラスターや表象というものを、いわゆる論理演算のように別のものを代入して **and** とか **or** をとって処理するといった形での、記号処理がどう行われているかは私もまだわからない。

■市瀬氏

人間だと、わりと抽象化して、一階述語、高階述語論理を考えられるので、そのような入れ子構造が、脳の中でどのように実現されているのかが、少しわかればヒントになるのかなと思う。

■山川氏

一階述語論理を扱うシステムが脳の中に元々入っていることはかなり、人間しかできないことから、なさそう。

人間はいろいろなプロセスを見まね、習得、学習できるので、単に学習できるプロセスの一つとして、そういうものがある。あまり脳のベーシックな仕組みの中に、そういった機能を持つといった仮定をする必要があるのかと、いったことに逆に疑問を持つ。

■銅谷氏

論理演算は、確率演算の特別な場合ととらえている。そう言った、確率モデルを生成する確率モデル、今回紹介した、[Tenenbaum グループの研究\(□7\)](#)は、ベイジアンプログラムラーニングによりそれができるとの話だ。

どれぐらい複雑な問題を実用的に解けるのかはわからないけれども、構造を一つに決めるのではなく、毎回深さや幅が確率的に変わってくるような、非常にフレキシブルな確率モデルを、想定することでいろいろな課題に対応しようとする。

原理的には入れ子構造も可能なはずだが、それが発散してうまく動かないことも考えられる。脳で確率モデルを生成する仕組みがあるとしたら、どういう風でしょう？何らかのリカレントネットワークを使ってできるかどうか。

■山川氏

それができるとしたら、脳のどの辺で処理されている可能性が高いか。

■銅谷氏

脳のどの場所で、という問題はあまり考えていないが、その原理の方が問題。

■会場から

モジュールの作られ方について、「ありうるタスクは大きい」とは、子供の意味のない行動のことを指しているか？

モジュールとどう関係するか、いろいろありうるタスクからだんだんせばまって、だんだん意味のある行動に集中したとき、モジュール的に分かれていくのか？生理モデルとしては研究が進んでいるのか？

■山川氏回答

「どのように脳のモジュールが作られるのか？」といった質問かとおもいます。一般的にタスクは、さらに小さなタスクを組み合わせとして表現できるような構造を持っているので、その構造に合わせて、何か脳の方もモジュールが作られていると思われます。一方で機能的にも階層性があり、ある能力は視覚に関わる機能だけかもしれませんが、他の機能がモジュール化されている場合もあるでしょう。一方でタスクに応じて組み合わせるべき能力が変化するという見方もできます。

■一杉氏回答

これは「様々なタスクがあり、それに対応したモジュールが世の中にあるのでは？」という質問でしょうか。それは、汎用人工知能の世界で言うビッグスイッチに相当していると思う。脳はそうでないというのが共通認識で、どんなタスクでも、脳のいろいろな部位が使われる。むしろ、タスクとモジュールは対応しないのではないのでしょうか。

■市瀬氏

タスクじゃなければ、ファンクションで別れているのか、それとも別のもので別れているのか？

■一杉氏

私としては、モジュールはいろいろあるが、聴覚野、言語野とか、脳の領野ごとに扱う情報は決まっている、扱う情報の種類によって分かれている。

■市瀬氏

ファンクションとして理解してよい？

■一杉氏

ファンクションではないですね。いろんなタスクで使われる情報。

■銅谷氏

それは情報の種類に応じたファンクションと見ていいのではないですか。

■山川氏

大変に議論が盛り上がってきたところ残念ですが、終了の時間が迫ってまいりましたがので、銅谷先生

【最後の一言】

■銅谷氏

全脳というときに、脳のいろんな場所がなぜちゃんとコミュニケーションできるのかの問題は非常に面白いけれども、あまりと研究されていないテーマと思う。

脳は一種のマルチエージェント系で、脳の各所に自律的に処理をするニューロンなり、局所回路がある。例えば頭頂葉は、視覚入力、体性感覚入力などいろいろな入力を受けるのだが、各入力線維にそれが何を表していますよというラベルが付いているわけではないのに、なぜかそれらをちゃんと統合して、有用な情報表現を作り出すことができている。

いわば通訳もいないいろんな国の言葉を聞きながら、なにか自分なりに理解することが、なぜかできてしまっている。こういったところの仕組みに迫る考え方といったものを是非作らなければならないのではと思う。簡単ではないかもしれないが、でも実際、脳はやっているわけだから、何かソリューションはあるわけで、それを理解していきたい。

■山川氏

銅谷氏には顧問をやっていただいております。

拍手でもって終了したいと思います。

【主な参考文献】

- (□1) Artificial Intelligence: A Modern Approach の Web サイト
<http://aima.cs.berkeley.edu/>
https://en.wikipedia.org/wiki/Artificial_Intelligence:_A_Modern_Approach
- (□2) 汎用人工知能アーキテクチャ CogPrime の資料
http://wiki.opencog.org/w/CogPrime_Overview
- (□3) ノーフリーランチ定理 [David H. Wolpert and William G. Macready 1995]
<http://delta.cs.cinvestav.mx/~ccoello/compevol/nfl.pdf>
- (□4) Generic Priors for Disentangling Factors of Variation
[Bengio 2013] <http://arxiv.org/abs/1305.0445>
Yoshua Bengio, Deep Learning of Representations: Looking Forward,
Statistical Language and Speech Processing,
Lecture Notes in Computer Science Volume 7978, 2013, pp 1-37
http://link.springer.com/chapter/10.1007%2F978-3-642-39593-2_1
- (□5) 線条体と中脳の間スパイラル構造 [Haber 2003]
Haber SN. , The primate basal ganglia: parallel and integrative networks.
J Chem Neuroanat. 2003 Dec;26(4):317-30.
- (□6) Koechlin グループの研究
http://annecollins.github.io/pdfs/papers/Donoso2014Science_Collins.pdf
- (□7) Tenenbaum グループの研究
<http://web.mit.edu/cocosci/Papers/Science-2015-Lake-1332-8.pdf>