

汎用人工知能の研究動向

Actuality in the field of artificial general intelligence

荒川 直哉^{*1}
Arakawa, Naoya山川 宏^{*2}
Yamakawa, Hiroshi市瀬 龍太郎^{*3}
Ichise, Ryutaro^{*1} フリーランス
Freelance^{*2} 富士通研究所
Fujitsu Laboratories Ltd.^{*3} 国立情報学研究所
National Institute of Informatics

This article gives an overview of issues, approaches and prospects for artificial general intelligence.

1. はじめに

本稿において「汎用人工知能 (Artificial General Intelligence, 以下 AGI)」とは、個別の課題に対して設計されるのではなく、さまざまなスキルを習得しうるように設計されるという意味で「汎用な」人工知能を指す。この発表では AGI へのさまざまなアプローチおよび実現の見通しについての概観を行う。

人工知能研究は、元来ヒトの知能に近い汎用的な機能を実現することを目指していた。また、専門分野外の人々が「人工知能」という言葉で想像するものも、ヒトの知能に近いものであろう。しかし、この目標を達成するのは当初想定したよりはるかに困難であった。

人工知能研究の創成期からの、いわゆる「古き良き AI (GOFAI)」では、記号処理によって汎用的な知能が実現できると想定していた(物理シンボル仮説)。早期に汎用的な人工知能を目指した SOAR [Laird 12] などの認知アーキテクチャ(後述)は、この仮説に基づいて設計された。このアプローチは現在も存続しているが、フレーム問題(後述)や非古典的なカテゴリーの問題⁴などの課題があることが指摘されている。また、知識がどのように獲得されるのかという学習の問題も解決されていない。

一方、非記号的あるいはサブシンボリックな計算を用いるコネクショニスト的なアプローチ(後述の機械学習やニューロコンピュータリングに対応する)は GOFAI と同じぐらい古い歴史を持つ。このアプローチでは学習に重点を置くが、記号や論理、言語の取り扱いに難があり、やはり未だ AGI の実現には至っていない。

AGI の実現が見通せないことから、人工知能研究の主流は個別の文脈において個別の「知的な」振る舞いをするシステムの作成を目指す「狭い AI (特化型 AI)」「カーツワイル 07] の研究へ移っていった。

21 世紀に入り、AGI への関心が再度高まってきている。2006 年には AGI の論文集 [Goertzel 06] が出版され、2008 年からは AGI 国際会議⁵が毎年開催されている。この背景には、電子計算機の計算能力の向上、脳科学や学習理論、ロボティクスの発展などがあると考えられる。

1 naoya.arakawa [at] nifty.com

2 ymkw [at] jp.fujitsu.com

3 ichise [at] nii.ac.jp

⁴ 例えば、多くのカテゴリーの包摂関係は、論理的に定義されるというより「プロトタイプ」からの距離に応じてフェジーに決定されるといったこと [レイコフ 93] [Gärdenfors 04] .⁵ <http://agi-conf.org/>

2. AGI の実現に向けた課題

AGI 実現に向けて、以下のような課題が指摘される。

2.1 フレーム問題

フレーム問題とは、デネットの定式化 [Dennett 84] によれば、行動選択の際に必要な関連する情報の検索および取捨選択に実際的でない時間がかかってしまうということである。これが物理シンボル仮説または類似の枠組みに固有な問題なのかどうかは明確でない。サブシンボリックな連想をベースとするシステムにおいては関連性の高い情報がまず連想されるため回避できるかもしれない。

2.2 記号接地(シンボルグラウンディング)問題

記号接地問題とは、記号がいかに実世界との関わりにおいて意味を持つかという問題である。つまりロボットなどの人工物において、直接センサなどを通じて得た外界の情報と、恣意的に意味づけされた記号との関連付けを成立させる問題に帰着する。

2.3 言語使用

言語使用は、[ピンカー 95] らが主張するように脳に生得的に組み込まれた機能だとすれば一つの特殊機能であるが、言語と論理の間には密接な関連があり [オールワード 79], 論理が汎用的なものであるなら言語も汎用的な能力だとも言える。言語使用の本質的な課題としては、記号的な人工知能にとっては自然言語が用いるカテゴリーの非古典性(脚注¹)が問題になり、サブシンボリックな人工知能にとっては言語体系の持つ生成性(有限の規則から無限のパターンを生み出せること)が問題になる [Arakawa 13]. 実際の言語処理の課題としては多義性解決が大きな課題である。

2.4 想像力

今ここにない過去、未来、他者といった想像された状況の表象(メンタルスペース)を操作する能力はヒトに特徴的な認知能力であると指摘されている [Fauconnier 03][松沢 11] . さらに、フォーコニエら(前掲書)は「想像」による「概念混合 (conceptual blending)」が人間の精神生活の本質であるという。また、想像する力が他人の心を理解する「心の理論」と関係することは明らかであろう。さらに関連して、汎用知能の一部分となるべき社会知能も、想像する力と言語能力に依存している。機械に人間並みの知能を持たせるためには、人間並みの想像力をいかにして与えるかということが課題になる。

3. AGI に挑む主なアプローチ

3.1 認知アーキテクチャ

人工知能分野においては、その創設期から(多くの場合は人間の)総合的な認知機能をモデル化する「認知アーキテクチャ」が研究・開発されてきた。これまでに上記 SOAR の他、ACT-R¹ など数多くの認知アーキテクチャが提案されてきている²。AGI もまた総合的な認知機能を実現しなければならないことから、上記の AGI 国際会議においても、OpenCog を始めとした認知アーキテクチャ関連の研究発表は多い。

認知アーキテクチャでしばしば重要視される機能であるプランニングは、現在状態から与えられた目標状態までの行動系列を計画する探索技術である。しかしながら大規模なプランにおいて、最初から詳細なレベルで計画を立てることはコストがかかるだけでなく、状況変化に対して脆弱になる弱点が生じてしまう。そこで実践的なプランニングとして、抽象的な意図(中間目標)を実行の必要性に応じて段階的に具体化する BDI アーキテクチャ [Rao 91] などが提案されている。

3.2 認知ロボティクス

AGI には、身体を与えられた時に人間と同様に環境の中で知的な問題解決を行えることや、人間の幼児と同様な知的発達を行うことが期待される。したがって、ロボットを用いた認知およびその発達の研究は AGI にとって重要であり、またロボットを用いた研究は AGI 研究のよいテストベッドとなると考えられる。さらに、具体的な身体を持つことは記号接地問題を克服する上でも重要であると考えられる。日本国内では例えば認知発達ロボティクスのプロジェクト^{3, 4, 5}や社会的知能発生学研究会⁶、さらに記号創発ロボティクス⁷の研究コミュニティがあり、海外では iCub⁸ のようなプロジェクトやコミュニケーションを学ぶロボットの研究 (VUB 研⁹での研究や Jido¹⁰など)がある。

3.3 機械学習

学習能力は AGI の重要な一部分である。なぜなら、同じ過ちを繰り返すシステムは知的とは言えないからである。近年、著しい進歩を見せる機械学習研究の主流は統計的な理論を駆使したサブシンボリックなアプローチである¹¹。機械学習理論の多くはベイズの確率理論をベースとする。一方、神経系にヒントを得たパーセプトロンなどを源流とするコネクショニストアプローチも統計的な学習理論に統合されてきている [ビショップ 12]。[Doya 06] や BESOM [一杉 08] のように脳そのものをベイズ理論で解釈するアプローチも現れてきている。

近年、大脳新皮質の階層と同様に多層化されたニューラルネットモデルを用いた深層学習 (deep learning) が、いくつかの領域のパターン認識タスクにおいて優秀な成績を収めている

([神鷹 13]による連載を参照)。人間並みのパターン認識は、人工知能研究が達成していない能力であるため、この分野での進歩は AGI の実現に向けた重要な進歩だと考えられる。

強化学習 [Sutton 00] も機械学習の一つであるが、認識される状態に対する価値付けを学習し、各状態においてその価値を高める行動を選択するというものである。そこで、エージェントは、ステップごとに報酬が与えられなくても試行錯誤的に報酬の期待値を最大化するような「ポリシー」を学習し、不確実な環境にも拡張されている。深層学習と強化学習の連携は最もホットな分野の一つである(例えば [Mnih 13])。

3.4 脳にインスパイアされた AGI

脳の知能には汎用性を含め、未だ電子計算機で実現できない計算機能がある。単に真似るだけで脳型計算機を作れるほどには脳は理解されていないが、AGI の有力なヒントにはなるだろう。脳にインスパイアされたアプローチには、脳全体をエミュレートする(真似る)認知アーキテクチャを目指す人工脳(Artificial brain)と、部分的な計算機能を機械学習として実現するニューロコンピューティング(NC)の二つがある。

人工脳¹²にはさらに、生物学的知見の再現を重視する大規模脳シミュレーション [Garis 10] と、計算機能を重視した Biologically Inspired Cognitive Architectures (BICAs) [Goertzel 10] がある。

大規模脳シミュレーションとしては、EPFL の Markram 氏による Blue Brain Project¹³、DARPA-IBM の Modha 氏らによる Cognitive Computation Project¹⁴、Stanford 大学の Boahen 氏による Neurogrid Project¹⁵ などがある。[Sandberg 08] によるロードマップによれば今世紀中頃には全脳エミュレーションの実現が予測されている。

BICA の研究では海馬もしくは大脳新皮質に関わるものが多い。身体性を重視した Edelman らの Brain-Based Devices [Fleischer 07]として、主に海馬機能を実装した移動ロボット NOMAD は有名である¹⁶。Milford らは齧歯類の海馬に学んだ自己位置推定と環境地図作成を同時に行う SLAM アルゴリズムを公開している [Ball 13]。[大森 04] は過去の経験を素早く再利用するメタ学習のための機能部品組合せ(FPC)モデルを提案した。また、一杉らは、機能的に分化した脳の各器官をできるだけ単純な機械学習器として解釈し、それら機械学習器を統合する全脳アーキテクチャ(WBA)の検討を開始している [一杉 14]¹⁷。

脳神経回路をヒントとして発展した機械学習技術としてニューロコンピューティング(NC)があるが、高次脳機能における計算処理と機械学習は、いずれもサブシンボリックな帰納学習であり、互いに相性がよい。近年は、不完全ながら脳の各器官の機能が特定されてきたため、興味のある計算理論を担う脳神経回路を特定し、その機構を参考として新たな計算方法を見つけ出そうとする研究アプローチが見えてきた。山川はこうしたアプローチを「理論にガイドされたニューロコンピューティング(TgNC)」と呼び、フ

¹ <http://act-r.psy.cmu.edu/>

² Comparative Table of Cognitive Architectures: <http://bicasociety.org/cogarch/architectures.htm>

³ <http://www.er.ams.eng.osaka-u.ac.jp/asadalab/index.html>

⁴ <http://www.isi.imi.i.u-tokyo.ac.jp/en/ninchi/ninchi.html>

⁵ <http://naotoiwahashi.jp/research.html>

⁶ <http://www.sociointelligenesis.org/>

⁷ <https://www.facebook.com/symbolemergence/>

⁸ <http://www.icub.org/>

⁹ <http://ai.vub.ac.be/> Luc Steels が主幹

¹⁰ <http://www.laas.fr/robots/jido/data/en/supervision.php>

¹¹ 朱鷺の杜: <http://ibisforest.org/>

¹² <http://www.artificialbrains.com/>

¹³ <http://bluebrain.epfl.ch/>

¹⁴ <http://www.research.ibm.com/cognitive-computing/neurosynaptic-chips.shtml>

¹⁵ <https://www.stanford.edu/group/brainsinsilicon/>

¹⁶ 1990 年台後半は RoboCup などの研究に利用された。

¹⁷ <https://staff.aist.go.jp/y-ichisugi/brain-archi/i-index.html>

レーム問題を担う基本的な計算機能が海馬にあると想定して、その脳部位での情報表現に着目しながら新たな計算アルゴリズムを見出そうとしている [Yamakawa 12].

大脳新皮質は多様な機能を複数の領野で実現しているが、領野を越えてほぼ様な神経回路を持つことが知られている [Mountcastle 82]. その神経回路の計算原理を解明できれば、AGI の実現への近道になるという期待が高く、人工脳や NC 研究で注目されている [ホーキンス 05] [George 09].

3.5 言語使用

機械の言語使用については、計算言語学や自然言語処理の分野で主に記号処理の側面から研究が続けられている。言語表現は一定の規則(文法)に従って理解され、生成される。規則に従った言語表現の機械的な処理は、多義性がなければそれほど難しくない。多義性の解消についても統計的な手法による研究が行われているが、照応などの文章構造の理解(例えば [Asher 03] を参照)を含めて考えると課題は多い。多義性の本質的な解消のためには、言語表現を「本当に理解」することが求められるのではないかと疑問がある。そこには非古典的なカテゴリの問題や、記号接地の問題が横たわっている。解決の糸口としては、レイコフら認知言語学者が主張するように身体性を重視すること、すなわち人工知能の文脈からするとロボットを使った研究が重要となるのではないかと考えられる。

3.6 普遍主義 AGI

上記 AGI 国際会議では概念的な議論も活発に行われているが、その1つに「普遍主義的 AI」と呼ばれるものがある [Hutter 05] [小林 14]. そこでは究極的な知能をオッカムの剃刀を数学的に定式化し、強化学習的な枠組みに適用した理論により定義している。この理論は計算不能で実際的ではないが、汎用知能の極北として議論され続けている。

4. 今後に向けて

4.1 ロードマップ

AGI の実現に向けてのロードマップを意図したものとしては、「人間レベルの汎用人工知能の実現への全体図制作」[アダムズ 14] (AI Magazine 2012 年春号)という論文においてヒトの発達段階をなぞることを想定したロードマップ(見取り図)が定義されている。上述の OpenCog プロジェクトでも認知発達に準拠したロードマップを設定している¹。

上記論文 [アダムズ 14] は、研究者間の共同作業の重要性を強調している。実際、ここまで述べてきたことがらは多くの専門にわたっている。すなわち、AGI の実現には、認知アーキテクチャ、ロボティクス、パターン認識、さまざまな学習理論、(認知・計算)言語学、認知心理学、脳科学などの知見を総合する必要があると考えられる。これらの分野の研究者が共同すべきことは当然であるが、個々の研究者にも互いの領域の基本知識を持つことが求められるだろう。

AGI のベンチマークとしては、古典的にはチューリングテストが提唱されており、具体的には毎年ローブナー賞コンテスト²が行われている。しかし、チューリングテストの基準は、文字ベースのチャットで人間と見分けがつかないことを持って人間を欺

くことであり、知能の汎用性の実現そのものを測るものではない。また、チューリングテストは、テキスト以外のモダリティや身体性についての能力については測ることができない。身体性を持った知能についてのベンチマークとしては Robocup に代表されるロボットのコンテストがある。特に、RoboCup@Home³では、人間とのインタラクションを含む複雑なタスクの実行が要求される。

4.2 実現時期の予測

カーツワイルは、情報科学の進歩は幾何級数的に起きること(収獲加速の法則)、脳科学は情報科学とみなせること、すでに大脳皮質の様な構造は解明されていることなどをもって、人工知能が十分な汎用性を獲得し、より高い知能を持つ人工物を自ら創造できるようになる時点、すなわち彼のいうシンギュラリティ(技術的特異点)は近いと預言している [Kurzweil 12]. しかし、彼は AGI 実現のための具体的なレンジを示しているわけではない。

2009 年から 2011 年の間に人工知能の専門家の間でいくつかのアンケート調査がなされた。それによると、専門家たちは AGI がこの数十年の間に起きるであろうと予想している。しかし、[Armstrong 12] によれば、専門家の予想はこの数十年にわたり相対的に数十年先の実現を予想している⁴。単なる期待以上の予測を行うのであれば、認知機能と実現手法を具体的に棚卸した上で、フィット&ギャップ分析を行ってみるべきだろう。

5. AGI 研究へのお誘い

5.1 関連コミュニティ

AGI に関連する国際的な動きとしては、上記 AGI 国際会議の他、BICA⁵ や Advances in Cognitive Systems⁶ のような会議が開催されている。国内においては、上記論文集の輪読会 [山川 14]、社会的知能発生学研究会⁷といった AGI に関連するコミュニティが活動している。関連研究を行っている方は国内外を問わず是非参加されたい。

5.2 日本における研究への期待

身体性は AGI にとって実現の手がかりとなる重要な要素であり、AGI を目指す研究にロボティクスを組み込むことにより、その実現を加速するものと考えられる。日本においては、認知ロボティクスの分野でも先進的な研究が行われている(上記)。また、上記記号接地問題との関係においては、記号創発ロボティクスの名のもとで研究が行われている [岩橋 12]。

また近年の神経科学知見の蓄積と機械学習の成熟、さらに大脳新皮質に似た構造の深層学習技術も進展を背景として、2013 年末より全脳アーキテクチャ勉強会⁸ が開催され、AGI を実現するため先に述べた全脳アーキテクチャ(WBA)と理論にガイドされたニューロコンピューティング(TgNC)の両面からの方法論が検討されている。

³ <http://www.robocupathome.org/>

⁴ cf. How long until human-level AI? Results from an expert assessment
http://sethbaum.com/ac/2011_AI-Experts.html
cf. <http://hplusmagazine.com/2011/09/16/how-long-till-agi-views-of-agi-11-conference-participants/>

⁵ <http://bicasociety.org/>

⁶ <http://www.cogsys.org/>

⁷ <http://www.sociointelligence.org/socio/>

⁸ 著者の山川までお問い合わせください。

¹ 上記参照 cf.

<http://wiki.opencog.org/wiki/home/images/3/39/Preschool.pdf>

<http://wiki.opencog.org/w/OpenCogPrime:Roadmap/>

<http://opencog.org/roadmap/>

² <http://www.loebner.net/Prizef/loebner-prize.html>

参考文献

- [アダムズ 14] アダムズ, S. et al.: 人間レベルの汎用人工知能の実現への全体図制作, 人工知能学会誌, Vol. 29, No. 3 (2014 in press)
- [オールウド 79] オールウド, J., アンデソン, L., ダール, Ö.: 日常言語の論理学, 産業図書 (1979)
- [Arakawa 13] Arakawa N.: Information Binding with Dynamic Associative Representations, *Formal Magic Workshop* (2013)
- [Armstrong 12] Armstrong, S. and Sotola, K.: How We're Predicting AI – or Failing To, in *Proc. Beyond AI: Artificial Dreams* (2012)
- [Asher 03] Asher, N. and Lascarides, A.: *Logics of Conversation*, Cambridge University Press (2003)
- [Ball 13] Ball, D., Heath, S., Wiles, J., Wyeth, G., Corke, P., and Milford, M.: OpenRatSLAM: an open source brain-based SLAM system, *Autonomous Robots*, Vol. 34, No. 3, pp. 149-176 (2013)
- [ビショップ 12] ビショップ, C. M.: パターン認識と機械学習 上/下, 丸善出版 (2012)
- [Dennett 84] Dennett, D.: Cognitive Wheels: The Frame Problem of AI, *The Philosophy of Artificial Intelligence*, Margaret A. B., Oxford University Press, pp. 147-170 (1984)
- [Doya 06] Doya, K., Ishii, S., Pouget, A. and Rao, R. eds.: *Bayesian Brain: Probabilistic Approaches to Neural Coding*, MIT Press (2006)
- [Fauconnier 03] Fauconnier, G. and Turner, M.: *The Way We Think: Conceptual Blending And The Mind's Hidden Complexities*, Basic Books (2003)
- [Fleischer 07] Fleischer, J., Gally, J., Edelman, G. and Krichmar, J.: Retrospective and prospective responses arising in a modeled hippocampus during maze navigation by a brain-based device, in *Proc. the National Academy of Sciences*, Vol. 104, No. 9, pp. 3556-3561 (2007)
- [Gärdenfors 04] Gärdenfors, P.: *Conceptual Spaces -- The Geometry of Thought*, MIT Press (2004)
- [Garis 10] de Garis, H., Chen S., Goertzel, B. and Lian R.: A world survey of artificial brain projects, Part I: Large-scale brain simulations, *Neurocomputing*, Vol. 74, No. 1, pp. 3-29 (2010)
- [George 09] George, D.: Towards a Mathematical Theory of Cortical Micro-circuits, *PLoS Computational Biology* (2009)
- [Goertzel 06] Goertzel, B. and Pennachin, C.: *Artificial General Intelligence*, Springer (2006)
- [Goertzel 10] Goertzel, B., Lian, R., Arel, I., de Garis, H. and Chen, S.: A world survey of artificial brain projects, Part II: Biologically inspired cognitive architectures, *Neurocomputing*, Vol. 74, No. 1, pp. 30-49 (2010)
- [ホーキンス 05] ホーキンス, J.: 考える脳 考えるコンピュータ, ランダムハウス講談社 (2005)
- [Hutter 05] Hutter, M.: *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*, Springer (2005)
- [一杉 08] 一杉 裕志: 脳の情報処理原理の解明状況, 産業技術総合研究所テクニカルレポート AIST07-J00012 (2008)
- [一杉 14] 一杉 裕志: 全脳アーキテクチャの解明を足がかりとした汎用人工知能の実現可能性, JSAI2014 OS-22 (2014)
- [岩橋 12] 岩橋 直人, 谷口 忠太: 特集「記号創発ロボティクスにあたって」, 人工知能学会誌, Vol. 27, No. 6, pp.544-545 (2012)
- [神寫 13] 神寫 敏弘, 松尾 豊: 連載解説「Deep Learning (深層学習)」にあたって, 人工知能学会誌 Vol. 28 No. 3, pp. 472-473 (2013)
- [小林 14] 小林 亮太, 相澤 彰子: 汎用エージェントの理論的枠組み, 人工知能学会誌, Vol. 29, No. 3 (2014 in press)
- [カーツワイル 07] カーツワイル, R.: ポスト・ヒューマン誕生—コンピュータが人類の知性を超えるとき (電子版: シンギュラリティは近い—人類が生命を超越するとき), 日本放送出版協会 (2007)
- [Kurzweil 12] Kurzweil, R.: *How to Create a Mind: The Secret of Human Thought Revealed*, Viking Adult (2012)
- [Laird 12] Laird, J. E.: *The Soar Cognitive Architecture*, MIT Press (2012)
- [レイコフ 93] レイコフ, G.: 認知意味論—言語から見た人間の心, 紀伊國屋書店 (1993)
- [松沢 11] 松沢哲郎: 想像するちから—チンパンジーが教えてくれた人間の心, 岩波書店, (2011)
- [Mnih 13] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. and Riedmiller, M.: Playing Atari with Deep Reinforcement Learning, in *NIPS Deep Learning Workshop* (2013)
- [Mountcastle 82] Mountcastle, V.: *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function*, MIT Press (1982)
- [大森 04] 大森 隆司, 小川 昭利, 山内 康一郎: 脳型と呼ばれる情報処理の定式化とその計算アーキテクチャ化を目指して: タスク認識に基づく即応的 FPC の試み, 信学技報 NLP, Vol. 104, No. 472, pp. 19-24 (2004)
- [ピンカー 95] ピンカー, S.: 言語を生み出す本能 上/下, 日本放送出版協会 (1995)
- [Rao 91] Rao, A. S. and Georgeff, M. P.: Modeling Rational Agents within a BDI-Architecture, in *Proc. International Conference on Principles of Knowledge Representation and Reasoning*, pp. 473-484 (1991)
- [Sandberg 08] Sandberg, A. and Bostrom, N.: Whole Brain Emulation: A Roadmap, Technical Report #2008- 3, Future of Humanity Institute, Oxford University (2008)
- [Sutton 00] Sutton, R. S. and Barto, A. G.: 強化学習, 三上 貞芳, 皆川 雅章 (訳), 森北出版 (2000)
- [Yamakawa 12] Yamakawa H.: Hippocampal formation mechanism will inspire frame generation for building an artificial general intelligence, in *Proc. AGI-12*, pp. 362-371 (2012)
- [山川 14] 山川 宏, 市瀬 龍太郎: 汎用人工知能輪読会の発足とその後の活動, 人工知能学会誌, Vol. 29, No. 3 (2014 in press)