

特集「汎用人工知能 (AGI) への招待」にあたって

山川 宏
(株式会社富士通研究所)

市瀬 龍太郎
(国立情報学研究所)

今回の特集テーマは「汎用人工知能」である。この言葉は **Artificial General Intelligence (AGI)** に対して、私達が選んだ和訳であるが、ここではあえて「人工一般知能」とは直訳しなかった。

最初のベン・ゲーツェル氏の記事で紹介されるように、AGI は「**Narrow AI**」に對置する概念として現れた。**Narrow AI** は特定の問題領域に特化した、すでに多く実用例をもつ **AI** の形である。対して **AGI** では、設計時には想定されていない広範囲の問題に対しても適応できる能力を、少なくとも人間程度にもつことを目指す未確立の技術である。そこで我々は、**AGI** の特徴は、一般というよりはむしろ汎用 (**General-purpose**) といったほうが似つかわしいと考え汎用人工知能と呼ぶことにした。

ところで汎用人工知能研究の文脈においては、常に人間レベルの知能という点が強調される。この背景として、計算機の発達した現代においては、**Narrow AI** が適用できる問題領域 (つまり、計算手段を機械学習の枠組みに分解し得る場合) では、しばしば人を凌駕しつつあるという現状がある。これはつまり、人工知能が広い問題領域に自律適応できる汎用性さえ獲得すれば、ほぼ全面的に人間の知能を超える可能性が高いという期待を示している。そして、その帰結として、人類は従来の傾向に基づいての技術進歩の予測が通用しなくなる技術的特異点 (テクノロジカルシンギュラリティ) を迎えることになり、その変化は社会に大きな影響を与える可能性が高い。

その一方で、汎用人工知能研究では、あくまで機械による高度な知能の実現を目指すというニュアンスも強く、“人間レベル”は目標としては本質的ではない。つまり、究極的に目指すべき汎用性は、人を超えていてもよいのである。そして人間レベルという目標設定はその評価についての意見が分かれやすい。そこでゲーツェル氏らは、汎用性という目標を掲げることでその困難さを緩和することを期待しているようである。

しかし現在の技術状況においては、汎用性については人間の脳が計算機をはるかに凌いでいる。そのため人間レベルの汎用人工知能という目的は、極めて実際的なも

のである。こうして、人間の心理的なプロセスや、その基盤となる神経科学的な知見、さらには生物進化などを含めた生物に学ぶ方略は、この目標に向けての有力な研究アプローチとなっている。近年、機械学習分野で話題となっている深層学習は脳新皮質を参考としており、ある意味でこうした流れをくんでいる。

特集の最初の記事は、この分野の世界的リーダーであるベン・ゲーツェル氏によるコンパクトでありながら包括的な汎用人工知能のレビュー記事である。**AGI** の中心仮説、特徴付け、諸観点、分野の範囲などについて語られている。

2 番目の記事では、ゲーツェル氏の記事内で普遍主義的アプローチと呼ばれる汎用エージェントの理論的枠組みを小林亮太氏と相澤彰子氏が紹介する。いかなる環境のもとでも最適な戦略をとるエージェントの定式化から理論モデル (**AIXI**) を展開する。

3 番目の記事では、荒川直哉氏とジェブカ・ラファウ氏が、昨年の汎用人工知能の代表的な国際会議である **AGI 2013** の会議報告を行う。この会議では認知アーキテクチャの研究が多く、それらの紹介とともに、基調講演、チュートリアルなどの様子を紹介している。

4 番目の記事は、篠田孝祐氏の監訳による和訳論文で、人間レベルの汎用人工知能実現に向けたランドマークの設定の試みなどを行う、壮大な記事である。元となった記事は、2012 年春に **AAAI** の **AI Magazine** に掲載された「**Mapping the Landscape of Human-Level Artificial General Intelligence**」である。知能の汎用性はある一つの課題のパフォーマンスによって見積もることができない。そこでこの記事では、人間の発達段階を踏まえた七つの典型的な評価シナリオが示される。

5 番目の記事では、「2045 年問題」の著者である松田卓也氏に、2045 年に技術的特異点が訪れるとした場合に、現在から 2030 年、2030 ~ 2045 年、2045 年以降と三つの時代に分けて、人類と社会において想定し得るシナリオについて論じていただいた。

6 番目の記事では、松原仁氏が、特定のタスクに特化して設計者がつくり込むことにより表面化しなかったフレーム問題と記号接地問題という二つの **AI** 基本問題

が、汎用人工知能においては重要になると指摘し、AGIに期待を込めつつも、個別の問題を着実に解いていくことの重要性も語られた。

最後の7番目の記事では、山川と市瀬が、本特集の母体となった汎用人工知能輪読会の活動状況を紹介して締めくくる。

本特集は、2013年の5月号の特集「シンギュラリティの時代」(Vol. 28, No. 3)に続く汎用人工知能とシンギュラリティに関わる2回目の特集である。しかしながら、この話題の範囲は広く、少なくとも以下の二つの点については、かなり不足があると考えている。

1点目として、汎用人工知能の実現を目指す研究として、ゲーツェル氏の記事で紹介されたような多くの認知アーキテクチャの研究が存在するが、それらの解説には及んでいない。

2点目として、AIが段階的に人レベルに近づくに従って社会への影響は、さらに拡大の一途をたどるものと思われるが、その範囲は、経済、政治、倫理、エンターテインメント、技術開発などの多岐にわたる。英国のAI開発企業DeepMind社^{*1}はGoogle社へ吸収合併される交換条件として、強力なAIが暴走したり悪用されたりしないようAI倫理委員会の設置を求めた例もある(2014年1月)。

本特集で不十分である点に関しても、我々としては、汎用人工知能輪読会を中心に、本学会全国大会のオーガナイズドセッションや、Webページなどを通じて、今後も情報を発信していきたいと考えている。

最後に、本特集の執筆者の皆様、さらに閲読などで協力いただいた汎用人工知能輪読会の皆様に改めて感謝する。

*1 共同創立者のShane Legg氏は、上記AIXIの共同提唱者である。