



来たるべきシンギュラリティと 超知能の驚異と脅威

応
般

1

松田卓也 (NPO あいんしゅたいん)

シンギュラリティとはなにか

シンギュラリティ (Technological Singularity, 技術的特異点) とは、人工知能の知的能力が全人類の知的能力の総和より大きくなること、またはその時点である。その時点より未来は予測が困難なので、一般相対論の時空のシンギュラリティにならってシンギュラリティと名付けられた。シンギュラリティという言葉は米国の数学者でSF作家であるV. Vingeが使い始めたものである。そこで生まれるであろう、人間よりはるかに高知能な存在を超知能 (Superintelligence) と呼ぶ。

シンギュラリティは人工知能の能力が爆発的に進化する現象、これを知能爆発 (Intelligence Explosion) というが、それと同一視することもできる。知能爆発の概念は1965年に英国の数学者I. J. Goodが提案した。彼は論文の中でこう述べている。「超知能とはいかなる賢い人をも遥かに凌ぐ知的な機械である。そのような機械の設計も知的活動なので、超知能はさらに知的な機械を設計できるだろう。こうして知能爆発が生じて、人類は置いて行かれる。超知能の発明が人類最期の発明である」

2014年に公開された米映画「トランセンデンス (Transcendence)」では、Johnny Depp扮する主人公の人工知能学者Will Caster博士は、シンギュラリティのことをトランセンデンス (超越) とよんでいる。

超知能が機械由来である場合、人工超知能 (Artificial Superintelligence)、人間由来である場合、超人類 (Trans Human) と呼ぶ場合もある。

シンギュラリティが起きる時期に関して、米国の未来学者Ray Kurzweilは2045年の頃としている

が、研究者によりいろいろな意見がある。本稿の目的は、超知能の種類、出現時期、取り組みの現状、人類社会への影響などについて論じるものである。本稿で論じる諸問題に関する解説として、Kurzweilの本¹⁾や筆者の本を参照されたい²⁾。

人工知能、超知能の分類

人工知能といってもさまざまある。以下でその分類を行う。

■ 狭い人工知能と汎用人工知能

人工知能は狭い人工知能 (Narrow Artificial Intelligence) と汎用人工知能 (Artificial General Intelligence : AGI) に分類できる。狭い人工知能は特定目的の人工知能であり、たとえばチェスや将棋をする人工知能、クイズ番組でチャンピオンを破ったIBMのWatson、それにiPhoneに搭載されているSiriなど、特定目的の人工知能である。現行の人工知能はすべて狭い人工知能である。つまり普通、人工知能といった場合はこちらを指している。

この狭い人工知能は特定目的では人間より遥かに強力である場合がある。たとえば前述のIBMの人工知能Watsonはジョパディ!というクイズ番組で人間のチャンピオンを破った。また将棋用の人工知能は、プロの棋士を破りつつある。今後、10～20年かけて発達するはずの人工知能は狭い人工知能で、人間生活に大きな影響を及ぼすであろう。

しかし本稿の主題は汎用人工知能である。汎用人工知能は本来の意味での「人工」知能であり、人間同様に意識、感情、常識をそなえたもので、それはまだ実現していない。

人工知能が意識を持つかどうかに関してはさまざまな意見があり得るが、持つとしてそれをどう検証するかが問題になる。伝統的な考えではチューリングテストにパスしたかどうかで判定する。チューリングテストとは被験者である人工知能と人間を、壁の向こうにおいて、試験官が彼らと会話をを行い、人間と人工知能の区別がつかない場合、人工知能に意識があると判定する。現状のテストの方法では、音声ではなく、テキストでの5分間の対話で、30%の試験官をだませれば良いとしている。その方法ですでにテストを通過した人工知能があるが、しかしそれでは不十分であるという意見が多い。前述の Kurzweil は、もっと徹底的なテストをした場合、コンピュータの能力の発展から推測して、人工知能が真のチューリングテストに合格するのは2029年のことだとしている。

■ 超知能の作り方

超知能とは現在の人間よりはるかに知能の高い存在と定義すると、それは必ずしも普通に考えられているような、コンピュータなどの機械による人工知能とは限らない。超知能実現の方法として、現在の人間の脳を薬や、脳に電流を流すなどして、知能を強化することも考えられる。しかし、それで現在の人間の知能を遥かに凌駕することは困難であろう。人間の脳の体積やニューロンの数に制限があるからだ。

生物的超人類

高い知能を持つ男女をかけ合わせて、さらに高い知能を持つ人類を生み出す方法がある。一種の品種改良である。この方法は直接行うには時間がかかりすぎる。一世代が最低でも20年程度かかるからである。何世代も掛け合わせるためには60～100年もかかるので実用的ではない。

そこで高知能の男女の精子と卵子を取り出して試験管ベイビーを作る。試験管の中で胎児ができたなら、その中から優れた遺伝子を持つ胎児を選択して、その胎児の細胞からさらに精子と卵子を作り、掛け合わせる。この方法では世代交代が短期間で行われる。

また別の提案としては、遺伝子工学の手法で遺伝

子を直接改造する方法がある。知能に関連した遺伝子は10,000ほどあり、それらを優れたものに置き換えると、知能指数1,000の人間が作れるという。実際、中国では世界中から優秀な知能を持つ科学者2,000人を選んで、その遺伝子を収集して、その解析を始めたという。

この種の手法は一種の優生計画であり、西欧ではキリスト教の伝統やナチスの優生計画を思わすので、実施は困難であろう。しかし宗教的、政治的制限が少ない中国では可能であろう。

知能増強

人間の脳とコンピュータを接続して、生身の人間の知能を強化する。これを知能増強 (Intelligence Amplification= IA) または知能補強 (Intelligence Augmentation) と呼ぶ。人間・コンピュータの共生系を作るのだ。

現在の人間もPCやスマホを用いることで、情報収集能力や計算能力に関しては、昔の人間より遥かに高知能であると言える。その意味でも、人間はすでにコンピュータと共生系を作っている。

これをさらに進めるためには、人間とコンピュータのインタフェースを改善して、人間とコンピュータの一体化をより進める必要がある。インタフェースの方法として、侵襲的方法と非侵襲的方法がある。侵襲的方法とは、肉体に手を加える方法で、たとえば脳にチップを埋めることも考えられる。非侵襲的方法は肉体に手を加えない、従来の方法である。

Kurzweilによると、赤血球程度の大きさのロボット、つまりナノロボットをつくり、大脳の血管に常駐させる。そして無線を通して、ニューロンの電位などを測定して外部のコンピュータに知らせる。あるいは外部の情報を、直接脳に伝える。こうするとキーボードやディスプレイなどを使わずに、脳とコンピュータが情報をやりとりできる。

人間を改造して作った人間とコンピュータの共生系はサイボーグである。つまり人間がサイボーグになることにより知能を強化する方法だ。

筆者はこの方法が、超知能の実現法としては好ましいと思う。ただし問題は、人間の知能だけ強化さ

れて、倫理性はそのままの可能性があることだ。その意味ではむしろ危険かもしれない。

集合知能

三人寄れば文殊の知恵というように、多数の人間を組織化して、組織としての知能を増強する。いわば衆知を集めるのである。これを集合知能 (Collective Intelligence) と呼ぶ。これは現在においても、会社などの組織で行われている方法である。組織は一個人にできないことができるという意味で、組織の知能は個人のものより高い。実際、現代の組織はコミュニケーション法の効率化により、昔の組織よりはるかに高知能になっているといえる。

しかしこの方法にも限度がある。たとえば大きな組織における官僚化の弊害などのために、思ったほど能率が上がらない可能性がある。またネットの議論に見られるように、衆知というよりはむしろ衆愚に陥る危険性もある。超知能実現の方法としては、困難性は少ないが、圧倒的な高知能を実現することは困難であろう。

人工脳による集合知能

遺伝子技術が進むと、体を持たない脳だけを幹細胞から作り出すこともできるであろう。その中から優秀な脳だけを集めて教育を施す。教育には肉体が必要だが、それはロボットないしはバーチャルな肉体で代替する。そうして育て上げた多数の脳からさらに優秀なものを選別して、それらを電線や光ケーブルで接続する。生身の人間ではなく人工脳による集合知能である。

この方法は脳だけの存在に人権があるかといった倫理的問題があり、実現には社会的抵抗があるであろう。

全脳エミュレーション

全脳エミュレーションとは死んだ人の脳を再現する方法である。死んだ人の脳をガラス化して、それをきわめて薄くスライスして脳の3次元構造を調べる、そして脳のニューロンとシナプスの結合状況のマップを作り、それにそっくりな構造を機械上で実現する。人間の脳が完全に機械で再現できるとしたら、スイッチを入れると、死んだ人の精神、魂が

よみがえる可能性がある。これは一種のマインドアップローディングであるが、後に述べる方法と異なり、死後に行う点が異なる。問題は若いときの聡明な脳ではなく、ボケた脳のコピーになるおそれがあることだ。

マインドアップローディング

人間の脳の働きを解明して、それが動くハードを作る。そして個人の意識をコンピュータに転送する。これをマインドアップローディングとよぶ。映画「トランセンデンス」で行われた方法である。

先に述べた知能増強は生身の人間が主体であり、人間が死ねば超知能も死ぬ。ここで述べるマインドアップローディングは、コンピュータに意識を転送するので、機械が主体になる。そのコンピュータにロボットを操縦させると、個人のアバターあるいはサロゲートができる。ロボットの肉体は必ずしも必要なく、バーチャルな世界で肉体を持たせることも可能である。

この種のマインドアップローディングの場合、生身の肉体が死んでも、人間の精神、魂は保存される。その意味では人間は、ある意味で不死になるのである。Kurzweil や後に述べる Dmitry Itskov が追求する方向性である。

機械人工知能

生物としての人間とはまったく別の人工知能である。現状のノイマン型コンピュータで脳の働きをシミュレートすることができるだろう。これは現在EUで進められているヒューマンブレインプロジェクトの方法である。

あるいはまったく違うタイプのコンピュータ、たとえばニューロモルフィックチップを使ったコンピュータを使って脳の働きをエミュレートできるかもしれない。あるいはノイマン型とニューロモルフィックチップでできたハイブリッドコンピュータを作り、その上で人工知能を実現する。ノイマン型コンピュータが理性を担い、ニューロモルフィックチップの方は感覚や感性を担う。この手法はIBMが米国国防省高等研究局の協力の下に行っているシナプス計画の手法である。

現状では、この種の機械知能が超知能実現の道として実現の可能性が最も高いであろう。

■ 超知能の3分類

ここでは別の観点による部類を行う。速い超知能、集合超知能、深い超知能である。コンピュータで例えれば、速い超知能とはクロック速度の速いもの、集合超知能とは並列処理である。

速い超知能

全脳エミュレーションでもコンピュータへのマインドアップローディングでも、コンピュータで脳をシミュレートした場合、そのクロック速度は人間と比べて圧倒的に速くすることができる。たとえば典型的なコンピュータチップのクロックサイクルを2GHz、人間のそれを200Hzとした場合、その比は 10^7 になる。

平均的な人間の知的能力を1ヒューマン(1H)と定義すると、コンピュータ上に実現される超知能の能力はこの場合 10^7 Hになる。速い超知能とは回転の速い頭脳ということになる。 10^7 Hの超知能が1秒で考えることを、人間なら 10^7 秒、つまり115日間、飲まず食わず、寝ずに考えたことに相当する。

並列処理の超知能

コンピュータ上に実現された人間並みの人工知能の能力が1Hであるとしよう。機械人工知能は人間とは異なりコピーが容易である。人間の誕生と教育に要する期間である約20年を待たずして、即座にコピーをつくることができる。それに要するハードがスーパーコンピュータ並みの能力を必要とするなら、コピーは予算的には簡単ではない。しかしそれが現状のPC程度の価格で実現できるなら、多数のコピーを作り、それを並列で動かすことにより超知能を実現できる。機械知能による集合知能である。

これは自分と同じ知能を持つ人間を、たとえば1,000万人集めた会社を作ることを考えれば良い。新しい飛行機を作るというプロジェクトを考えると、それを多数の部分的な仕事に分割して、多数の人が共同作業を行う。これと同等のことを1台の集合超知能が行うのである。

速い超知能は、時分割により集合超知能と同じことができる。つまり速い超知能は集合超知能を包含する。仕事には並列処理できないものも存在するので、速い超知能のほうが、集合超知能より優れている。コンピュータで言えば、速いクロックのコンピュータのほうが、同等の並列コンピュータより優れているということである。

深い超知能

上述の超知能は普通の人間の知能を、クロック速度を上げたり、多数集めたりするもので、その考えは深くはない。天才的な数学者なら解ける問題は、凡人は一生かかっても解けないだろう。あるいは凡人を1,000万人集めても解けないだろう。天才的な数学者の考えが深いからである。鉄砲はいくら集めても大砲にならない。射程距離が違うのである。

深い思考ができる超知能を作る方法は明らかではないが、ハードをより複雑にすることで達成可能かもしれない。Giulio Tononiの統合情報理論によれば、意識の深さはシステムの複雑さによるという。人間の脳のニューロンの数は 10^{11} 程度、1つのニューロンには 10^4 程度のシナプスがあり、他のニューロンとつながっている。機械知能は、たとえばこの数を増やす、人工脳なら脳の体積を増やすなどの方法により、より深い知能を持つ超知能が作れるかもしれない。

具体的な計画

前章では超知能のさまざまな作り方に関して議論した。本章では現在進行形の、あるいは近い将来に行われるであろう、さまざまなプロジェクトについて紹介する。

■ ヒューマンブレインプロジェクト

ヒューマンブレインプロジェクト(Human Brain Project)とはEUが2013年から10年間かけて、総額12億ユーロ(1,600億円)ほどを投じて脳を解明するプロジェクトである。90の研究機関が参加している。

さまざまなターゲットがあるが、主要なものとしてスーパーコンピュータで脳を第一原理からシミュレーションしようとするものがある。スイスのローザンヌにあるスイス連邦工科大学の Henry Markram が主体となって行っている計画である。Markram はすでにヒューマンブレインプロジェクトの前身として、2005年に始まったブルーブレインプロジェクトを成功裏に終えている。それは IBM との共同研究で、IBM のスーパーコンピュータ、ブルージーンを用いて行われた。

大脳の構成要素である、10万個のニューロンからなる新皮質コラムの中で起きる現象を、化学反応のレベルからシミュレートしようという方法である。Markram はすでにネズミ程度の知能は実現できたとしている。つぎに猫程度の知能、さらに猿程度と進んで、2023年には人間程度の知能を実現するとしている。

Markram の手法は次に述べる IBM のシナプス計画と異なり、化学反応のレベルから大脳をシミュレーションすることにより、意識を浮かび上がらせようというものである。ボトムアップのアプローチだ。Markram は IBM のシナプスの方法論はナンセンスであり、その報道された成果はサギであるなどと厳しく批判していた。

一方その Markram の手法にも研究者の間には懐疑論があり、また予算の配分を巡ってプロジェクトメンバから厳しい批判が起きている。ヒューマンブレインプロジェクトが脳のコンピュータシミュレーションに主眼をおき、脳・神経科学の実験的研究への配分を減らしたからである。脳・神経研究者と人工知能研究者の予算を巡る争いであろう。

頭脳を一から解き明かして、あわよくば超知能を作ろうという Markram の計画は、あまりに大胆すぎて、一部の研究者には受け入れられない。はたして2023年までに人間の脳と同じ働きをするシミュレーションが可能かどうか、これからの成果を見なければならぬ。

■ シナプス計画

米国 IBM のシナプス計画 (SyNAPSE: The Systems of Neuromorphic Adaptive Plastic Scalable Electronics: 神経形態学的電子工学システム) は従来のフォンノイマンアーキテクチャに基づくコンピュータチップとはまったく異なるチップを作ろうという野心的な計画である。プロジェクトは2008年に始まり、米国国防省高等研究計画局から約5,300万ドルの資金援助を受けて行われてきた。すでに第0, 1, 2, 3段階は終わり、2013年にはトゥルーノースと名付けられた新しいチップを発表した。研究グループの主導者は Dharmendra S. Modha である。

この新チップは100万個のプログラム可能なニューロンと、2億5,600万個のプログラム可能なシナプスから構成され、Samsung の28ナノメートルプロセス技術で製造され、54億個のトランジスタからできている。消費電力は70ミリワットで、これは通常のCPUチップに比べると圧倒的に小さい。大きさは切手程度である。IBM はこれらのチップを組み合わせて、1枚のボードに1,600万個のニューロンと40億のシナプスからなる16チップを載せたシステムを公開した。

ちなみに人間の頭脳には、大脳で数百億、脳全体では1,000億個のニューロンがある。1つのニューロンには1万個に及ぶシナプスがある。ニューロンの数はマウスでは7,000万個、ラットで2億個というから、ニューロンの数だけから言えば、後一步でマウスに近づく。

このチップでニューラルネットワークを形成するが、内部パラメーターを自立的に決定する学習の機能はない。学習は外部のスーパーコンピュータを用いて行う。物体認識、音声認識、数字認識、視線認識などに適している。

IBM はこのニューロモルフィックチップを使った計算を認知計算 (Cognitive Computing) と呼んでいる。Modha は従来のノイマン型コンピュータを左脳型、ニューロモルフィックチップを使ったコンピュータを右脳型とよび、両者を合わせることによ

り、完全な人間の知能が再現できるとしている。

ニューロモルフィックチップは従来の CPU とは原理が異なるまったく新しい CPU であるので、そのソフトウェアはまだ成熟していない。IBM はすでに新チップ用の言語を開発し、さまざまなライブラリを公開している。そして外部の大学や研究所にソフトウェアの共同開発を呼びかけている。

化学反応のレベルから脳の活動を解明しようというヒューマンブレインプロジェクトは、例えてみれば空を飛ぶのに、鳥を研究して、鳥そっくりなものを作ろうというようなものだ。それに対してシナプス計画の方法論は、鳥の飛行を研究するが、飛行の実現法としては、鳥とは異なる固定翼を採用する飛行機を作ることに例えられるだろう。実際の動物のニューロンの活動はアナログ的であるが、新チップはデジタル的である。もっとも他のグループの開発しているニューロモルフィックチップにはアナログ的なものも存在する。

■ Watson

Watson は IBM が開発した質問応答システムで、2011 年に米国の人気クイズ番組の「ジヨパディ！」で人間のチャンピオン 2 人を破ったことで脚光を浴びた。IBM はすでに 1997 年に当時の世界チェスチャンピオンの Garry Kimovich Kasparov をスーパーコンピュータのディープブルーで破ったが、それにつぐ人工知能プロジェクトである。

2011 年当時の Watson のハードは 10 台のラックに搭載されたパワーシステム 750 で 3.5GHz で 8 コアのパワー 7 プロセッサを 2,880 個搭載していた。コアごとに 4 スレッド処理できる。メモリは 16TB で処理能力は 80TFLOPS であった。OS は Linux である。ハードは 2014 年にはさらに 90% 小さくなり、速度は 24 倍速くなっている。

ジヨパディ! 挑戦当時の Watson はインターネットには接続されておらず、本、百科事典、ウィキペディアなどを含む 2 億ページ分のテキストデータを持っていた。応答は自然言語で行われた。入力テキストを入力し、出力は音声で行われた。

Watson はその後、医療分野でがんの診断などのテストに使われている。2014 年に IBM は 2,000 人の人材と 1,000 億円を投資して Watson を商用に応用するビジネスを開始した。目的は薬品開発の研究支援、ビッグデータの可視化、自然言語を使用した質問への解答などである。

■ ブレイン計画

米国の Obama 大統領は 2013 年の一般教書演説で脳活動のマッピングに関する研究プロジェクトであるブレイン計画 (BRAIN Initiative: Brain Research through Advancing Innovative Neurotechnologies) を立ち上げることを宣言して、2014 年度に 1 億ドルの政府予算を提案した。

このプロジェクトの目的は脳の活動を電極による記録と、脳機能画像のレベルの間を埋めるナノテクノロジーを開発すること、脳のある領域のすべての活動を記録する技術を開発すること、単純なモデル動物から始めて神経回路の解析を行い最終的には人の脳機能を解明すること、というものである。

このプロジェクトはヒューマンブレインプロジェクトに触発されて始められたと思う。人間の脳は、宇宙の進化の中で達成された最も複雑なものであり、その解明は宇宙の誕生、生命の誕生の謎の解明と並んで、21 世紀の最大の科学探究目標であろう。

■ アバタープロジェクト

ロシアの企業家 Dmitry Itskov は 2045 年までにシンギュラリティを起こすための活動を開始した。彼の究極的目標は死の克服である。2013 年に「世界の未来 2045」という国際会議をニューヨークで開催して、シンギュラリティに関係した世界の専門家、著名人を招待した。日本からも、人間そっくりなアンドロイドで有名な阪大の石黒浩教授が参加した。Dalai Lama も Itskov の考えを支持しているという。

Itskov は彼の目標を達成するためにアバター計画を立案した。それは A ~ D の 4 段階からなっている。アバター A では 2015 ~ 2020 年に脳でコントロー

ルできるロボットを作る。Bでは2020～2025年に脳を、合成された肉体に移植する。Cでは2030～2035年にマインドアップローディングを達成する。Dでは2040～2045年に肉体の代わりにホログラム人間を作りだす。

Itskovは国際会議開催だけでなく、私費を投じて研究所を作り、研究者を雇って研究をしている。Itskovの野心的な計画が実を結ぶのか、あるいは単なる夢に終わるのかは、まだしばらく様子を見なければならぬ。

■ 米国 IT 大企業の取り組み

企業レベルで人工知能、ロボット研究のトップを走っているのは Google である。実際、Google の取り組みは予算の面から見ても、政府レベルの研究を遥かに凌駕している。Google は自社開発のほか、他の会社の吸収合併に膨大な資金を投じている。今までの10数年間で Google は M&A に年平均1,500億円ほど投じてきた。最近はそのペースが加速している。

2013年から2014年にかけての Google の動きは凄まじい。Boston Dynamics 社を含むロボットの会社8社を買収した。そのなかには東大発のベンチャー企業も含まれている。人工知能入りの火災報知器を作っている会社 NEST を数千億円で買収した。さらに人工衛星、監視カメラの会社も買った。全世界のデータを収集するのが目的であろう。

Google は英国の人工知能の会社 DeepMind 社を数百億円で買収した。この会社は人工知能の最新理論であるディープラーニングの研究者を多く抱えている。しかしこの会社はまだ製品を発表していないのだ。その創始者が天才的な人物であるので、創始者と社員の才能を買ったのだろう。

天才と言えば、Google は Kurzweil を雇って自社の研究所所長に据えた。Kurzweil は意識の核が自然言語理解にあると考えている。Google はさらにディープラーニングの創始者の Geoffrey Hinton を雇った。これからの世界は頭脳の勝負である。天才をどれだけ集めるかで勝負が決まる。

Google は世界初の商業用量子コンピュータ D-Wave を買って、NASA の構内に量子人工知能研究所を設立した。さらに Google は D-Wave の方式に習った量子コンピュータを自社開発しようとしている。

他の会社も負けてはいない。Facebook は DeepMind 社の買収合戦では、Google に破れた。そこで仮想現実用のヘッドマウントディスプレイを開発した Oculus VR 社を数千億円で買収した。

Microsoft も最近では人工知能に投資して自動翻訳システムを発表した。Amazon はロボットの Kiva Systems を買収したし、巨大なコンピュータセンターを運営している。

これらの米国の民間企業の競争を見ていると、その投じる予算は米政府や EU より遥かに巨額である。国家機関では意思決定は何重もの委員会の議論を通らなくてはならない。しかし Google のような民間企業ではトップがやると言えば、できるのである。超知能を初めて作り出すのは、政府の予算で運営される研究機関ではなく、Google のような民間企業ではないかと思う。

中国では百度という検索会社がシリコンバレーに人工知能研究所を作り、スタンフォード大学の人工知能の権威 Andrew Ng 教授を雇った。

今後の世界は政府ではなく、多国籍大企業に支配されるのではないだろうか。

■ 日本の取り組み

それでは日本の取り組みはどうだろうか。筆者の知る限り、政府と大企業の大規模な取り組みはゼロである。個々の研究者、研究グループの、1年で数億円規模の人工知能研究、脳・神経研究はあるだろうが、上述したような政府レベル、大企業レベルの数百億～数千億円に上る大規模な研究計画は聞いていない。

日本政府は1980年代に第5世代コンピュータ計画という野心的な人工知能計画を世界に先駆けて行った。しかし結果的には大失敗であった。大きな予算を投じた割には、論文以外はなにも有用な製品が

できなかったのである。多分それに懲りたのであろう、研究者の個人レベルの細々とした研究以外は、日本政府も大企業も超知能にはまったく興味を示していないように見える。

■ 全脳アーキテクチャ勉強会

それでは日本にまったく目がないかというところでもない。まったくの手弁当であるが、日本からシンギ

ュラリティを起こそうという運動が産総研の一杉、富士通の山川、東大の松尾たち若手研究者を中心として2013年に始まった。「全脳アーキテクチャ勉強会」といって、月一度程度の勉強会を行っている。

大脳アーキテクチャ勉強会の発起人の1人である一杉によれば、脳に関してはすでに膨大な知見があり、脳は普通の情報処理装置であり、それほど複雑な組織でもなく、計算量的にもシミュレーション可能であるという。ここ数十年で脳の理解のために必要な知識、たとえば機械学習や神経科学の知見が成熟した。

具体的には言語理解を司る大脳皮質のモデルとしてはベイジアンネットワークやディープラーニング、大脳基底核は強化学習、海馬は自己連想ネットワーク、滑らかな運動を司る小脳は教師あり学習、情動を司る扁桃体は連合学習でモデル化可能だと言う。

山川によれば脳をモデルとする全脳アーキテクチャ的方法論はディープラーニング+があれば必要な部品はそろっているので、20年代前半にも、シード（種）となる汎用人工知能の実現は可能だと見る。

この辺りの話は個人の見解であるから、まだ何とも言えないが、超知能の実現が海のものとも山のものともつかないといった状況ではない。そうだとすれば、日本でも欧米諸国のように国家の総力を挙げて研究開発を行えば、日本からシンギュラリティを

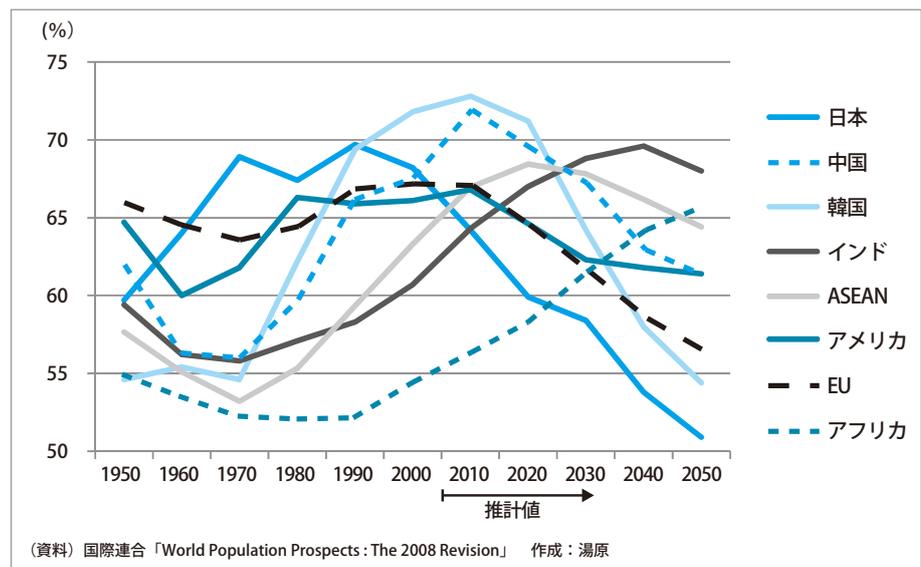


図-1 生産年齢人口割合の時間変化の各国比較

実現することは可能かもしれない。そのためには優秀な人材を1カ所に集めた大々的な研究機関を作ることである。

次の章でシンギュラリティ、超知能がなぜ日本で必要なのか、その政治的、社会的意味について論じる。

超知能の政治的、社会的意味

■ 超知能と日本の将来

日本は1960～70年代に高度経済成長を遂げて、1980年代にはジャパンアズナンバーワンと言われるまでになった。ところが1990年代に入り突然、失われた20年といわれる停滞期に入った。そして2011年の東日本大震災と福島原発事故を契機とするかのように、急速な衰退を始めた。その原因の1つは原子炉の運転停止による化石燃料の輸入の増加と、それによる国富の流出であろう。しかし日本の停滞はすでに90年代から始まっているので、根本的な原因は少子高齢化による生産年齢人口割合（労働力人口比）の減少であろう（図-1参照）。つまり国民に占める労働者の数が減少し続けているのが問題である。

生産量は労働者数と生産性の積であるので、労働者人口が減っても生産性を上げれば、取り戻せるはずである。その生産性を上げるのが人工知能化、口

ポット化である。超知能を開発することができて、それを制御することができたとすれば、ほとんど無限の生産性向上が見込める。

超知能の出現は危険かもしれないが、しかし先に述べたように、日本がやらなくても、米国はやるのである。このまま座して滅びるのを待つか、一か八にかけるか、日本の指導層の決意にかかっているのである。しかし残念なことに日本の指導層は、このことにまったく気づいていないように見える。

■ 超知能と核の類似性

筆者は超知能とは核のような物だと考えている。諸刃の剣なのである。核は平和利用すればエネルギーを生み出し、富を創出することができる。しかし核兵器となると、人類を滅ぼす可能性もある。

そんな危険な物をなぜ開発したのか。それは米国がドイツの核開発を恐れたからである。原子爆弾を開発したとき、一部の科学者は、核爆発を起こすと、地球の空気がすべて連鎖反応を起こし爆発してしまう危険性を警告した（もちろん現在の知識では、そんなことは起きないことは分かっているが）。しかし米国の軍部は核実験を強行した。やらなければやられると思ったからだろう。

超知能も同じことだ。うまく利用できれば、無限の富を生むことができる。アメリカのIT大企業である Google, Apple, Facebook, Amazon, IBM, Microsoft などは、最近、膨大な投資を人工知能とロボットに対して行っている。その目的は一言で言えば、金を儲けるためである。

超知能は軍事利用すれば、敵対国に対して絶対的な優位性を獲得できる。アメリカがやらなければ、中国やロシアがやるだろう。米国国防省高等研究局は人工知能とロボットの開発に多大な投資をしている。アメリカでは国家安全保障という言葉は魔法の言葉で、何でもできてしまうのだ。

つまり金儲けと防衛、標語的に言えばウォールストリートとペンタゴンがある限り、人工知能開発は止められない。

■ 超知能と世界覇権

歴史を振り返ると 20 世紀前半までの、世界覇権のための武器は戦艦であった。日本は日露戦争において戦艦でロシアに打ち勝った。第一次大戦でも戦艦は大きな役割を果たした。その後、米英日は世界覇権国家になり、戦艦の建艦競争を繰り広げた。しかし第二次大戦において、戦艦ではなく航空機の優位性が明らかになった。原子爆弾もそうである。20 世紀後半には、米ソを中心とする大国は、核兵器、航空機、ロケットの開発競争を繰り広げた。筆者の意見では、21 世紀の世界覇権を握る武器はもはや核兵器ではない。コンピュータとネット、人工知能とロボットである。それは単に軍事的な意味だけではなく、経済的、科学技術的、文化的覇権を握る武器なのだ。

超知能と核の類似性について述べたので核開発の歴史を少し振り返ってみよう。19 世紀の終わり頃に Antoine Henri Becquerel や Maria Skłodowska-Curie (キュリー夫人) が放射能を発見した。20 世紀に入り、量子力学や原子物理学、原子核物理学が急速に進んだ。核に関して言えば 1938 年のウラン核分裂の発見が大きい。その発見を受けて、アメリカでは国家の総力を挙げて核開発計画であるマンハッタン計画を行った。1942 年にはシカゴ大学で初の原子炉ができた。その 3 年後の 1945 年には広島と長崎で原子爆弾が爆発して、日本は降伏した。指導層がその気になった核開発の速さは驚異的である。

もし Kurzweil の予測が正しいとすれば、日本敗戦の年である 1945 年のちょうど 100 年後にあたる 2045 年には、原子爆弾の爆発ではなく、知能爆発のために日本は再び敗れるだろう。原子爆弾の開発に完全に遅れをとったように、超知能の開発にも、まったく遅れをとっているからだ。

シンギュラリティに関する諸問題

■ シングュラリティの起きる時期

人工知能学者の多くはシンギュラリティが 21 世紀中に起きるだろうと予測している。先に紹介した

Kurzweil は、チューリングテストをパスする人工知能が 2029 年、超知能が 2045 年に実現すると予測している。

人工知能学者のいくつかの世論調査では、シンギュラリティが発生する、つまり超知能が誕生する時点として、2020 年代前半とする人が 10% 程度、40～50 年とする人が 50%、21 世紀中とする人が 90% であるという。もちろん、そんなことは起きないとする人も 10% はいる。

人間にとって重要なことは、超知能ができたときにどうなるかである。つまり人間が超知能を使いこなすのか、あるいは滅ぼされるのか。これも専門家の意見分布だが、超知能ができることは人類にとって 1) きわめて良いことだ: 20%, 2) まあ良いことだ: 40%, 3) 中立: 15%, 4) かなり良くない: 15%, 5) きわめて悪い: 10% である。楽観論が多い。

Kurzweil は楽観論で、人類がマインドアップロードして超人類になることにより、人類のほとんどの不幸の源泉は解消すると主張している。たとえば人間が不死になるというのだ。

逆に、超知能ができると人類は滅ぼされる可能性があるのだ。作るべきでないと主張する人もいる。それについては後で議論する。

■ 離陸期間

超知能に危険性があるとして、安全策を取るには時間の余裕が必要である。人間並みの人工知能はそれほど危険ではないだろうが、人間の知能を遥かに凌駕する超知能は危険かもしれない。人間並みの人工知能ができてから、超知能に進化することを離陸と呼ぶ。離陸の期間がどのくらいに関して 3 つの可能性に分類できる、

- 1) 速い離陸: 1 分, 1 時間, 1 日
- 2) 中間的離陸: 数年
- 3) 遅い離陸: 10 年以上

Kurzweil はムーアの法則に基づいて、離陸期間を 15 年程度と想定しているのだから遅い離陸に分類できる。

しかし、Good がいうように、一度人間並みの強

い人工知能（シード AI）ができると、自分を改良して、進歩が加速する可能性がある。人間並みの人工知能ができたとすると、そのコピーを作る。そのコピーがスーパーコンピュータでしか走らないとすると高価なので、おいそれとはできない。しかし PC 程度で十分とすると、多数のコピーを容易に作ることができる。そして多数の人工知能に、自分自身の改良の研究を並行してさせる。すると以前の物より少し賢い人工知能ができるだろう。そのコピーをまた多数作り、次世代の人工知能を作らせる。このようにして人工知能は代を経るごとに賢くなっていく、この場合、知能爆発が起きて、速い離陸となる可能性が高い。速い離陸の場合、気がついたときには、もはや制御不能になっている可能性がある。

■ 超知能は制御できるか

超知能は人間より遥かに知能が高いので、もしそれを制御できるか、あるいは人間のために働いてくれるなら、ほとんど無限の富が創出できるだろう。人間は働く必要がなく、一生遊んで暮らせるかもしれない。

しかし人間より遥かに賢い存在が、人間の言うことを聞いてくれる保証はない。人間は動物を飼っているが、生殺与奪の権は人間が握っている。猫や犬には愛情を注ぐが、牛や豚、鶏は殺して食べてしまう。現状では人間は食物連鎖のトップにいるのだが、超知能が出現すれば、トップの地位が機械にとって代わられるかもしれない。人類が超知能に飼いなされるか、取って食われるか、それは分からない。人類が超知能を飼いならすことはきわめて困難であろう。犬あるいはゴキブリが人間を飼う状況を想像すれば良い。

James Barrat は人間に友好的な人工知能 (Friendly AI) が可能か、超知能が人間に対して反乱しないように閉じ込めておけるかなどの問題、超知能の安全性に関して専門家にインタビューした結果、それは不可能きわめて困難であると結論して、超知能研究の中止を求めている。

Louis A. Del Monte はシンギュラリティを研究し

て、21世紀後半には、機械由来の超知能と人間由来の超知能が共存するが、最終的には機械由来の超知能が世界を支配するだろうと予測している。

■ de Garis の超知性戦争

オーストラリア出身の人工知能学者 Hugo de Garis は21世紀後半には、人類の知能の1兆の1兆倍 (10^{24}) 倍の知能を持つきわめて知的な神のような機械ができるとして、それを超知性 (Artilect=Artificial Intellect) と名付けた。

上の数字の根拠は次のようなものだ。その時代になれば1原子で1ビットの情報を蓄えることができるだろう。1モルの物質 (炭素なら12グラム) にはアボガドロ数、ほぼ 10^{24} 個の原子がある。それが1フェムト秒 (10^{-15} 秒) でスイッチングすると仮定する。すると10モルの原子で1秒あたり 10^{40} 回スイッチングできる。

一方、人間の脳には 10^{11} 個のニューロンがあり、それが 10^4 個のシナプスを持ち、10Hz でスイッチングできると仮定すると、1秒あたり 10^{16} 回スイッチングできる。その比は 10^{24} である。ちなみに京コンピュータの速度は 10^{16} FLOPS 程度であるから、人間の脳と同程度の情報処理能力を持っている。

de Garis はそのような超知性ができる人間は滅ぼされるだろうと予測する。そこでそのような機械を作るべきかどうかで、人々の間に大論争が持ち上がるだろうという。超知性を作るべきとする一派を

de Garis は宇宙派、作るべきでないとする一派を地球派と名付けた。そして両派の間に大戦争が起きるだろうと予測して、それを超知性戦争 (Artilect War) と名付けた。

de Garis の予想 (妄想?) はさらに飛躍する。神のような超知性は、真空のゆらぎから新しい宇宙を作り出すだろう。そしてその宇宙からまた人類のような知的生命が現れ、それが神を作る。キリスト教の教えではまず神があり、それが世界を作り、人間を作ったとしている。de Garis は逆だという。人間が神を作り、その神が宇宙を作るという。ここまで来ると完全にSFの世界であるが、その予想される実現時期が千年後とか来世紀ではなく、今世紀であるというところが、ワクワクするし、また恐ろしいのである。

参考文献

- 1) レイ・カーツワイル：シンギュラリティは近い—人類が生命を超越するとき、Kindle版、NHK出版 (2010)；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、日本放送協会 (2007)。
- 2) 松田卓也：2045年問題—コンピュータが人類を超えるとき、廣済堂出版 (2013)。

(2014年10月6日受付)

松田卓也 | tmatsuda312@gmail.com

1970年京都大学理学研究科物理学第二博士課程修了、理学博士。神戸大学名誉教授。現在NPO法人あいんしゅたいん副理事長、ジャパン・スケプティックス会長。著書に「間違いだらけの物理学」学研 (2014)。