1.5 ビッグデータ時代の知識処理

1.5.1 総論

図26　本節の構成

人工知能（AI）はデータや知識を参照して有意な結論を導くが、データや知識の形態と量、またそれらから結論を導く方法（推論法とも言える）は時代とともに変遷がある。

第2次AIブームであった1980年代の知識処理は、主にルールで記述された知識に基づき、演繹を主体とする推論により結論を導くものであり、多くのエキスパートシステムが作成された。しかし、知識の記述は人手によるものが大半であるために大規模化するのが困難で、データ／知識の量は今日のビッグデータ時代に比べるとはるかに小規模であり、期待には応えられなかった。この人手による知識記述や知識獲得の課題に対処すべく、1980年代後半から機械学習、データマイニングの研究が活発になり、今日に至っている。

1990年代に普及、拡大したウェブは、グローバルな情報共有、情報流通のプラットフォームとして情報環境に革命的な変化をもたらし、今日に至っている。知識の点では、共同執筆のオンライン百科事典である「Wikipedia」は、オンラインの知識源として重要な役割を果たすようになっている。また2000年代には「SNS」（Social Network Services）の登場等で、各個人の情報発信が増大した。このようなサイバー情報空間の拡大に加え、「IoT」（Internet of Things）による各種センサからのデータもサイバー空間にもたらされ、サイバーフィジカルシステム（Cyber Physical System; CPS）も進行しつつある。このようにしてデータ／知識量は増大しつつあり、ビッグデータ時代となった。

このビッグデータ時代のデータや知識の形態は、1980年代とは異なるものであり、それらを利用する知識処理も変化してきている。深層ニューラルネットワーク（DNN）によるディープラーニングは、大量データに基づき高性能な識別性能を達成する機械学習法として、2012年以降、広く利用されるようになってきている。

本節ではまず、以上のようなビッグデータ時代のAIが基にするデータや知識とその利用法を概観する（1.5.2項参照）。次いで、「データのウェブ」ともいうべき、知識をコンピュータが意味把握しやすい規格化された形態で表すLOD（Linked Open Data）の動向と、それに関連するオントロジー（基本語彙体系）について紹介する（1.5.3項参照）。最後に、ビッグデータを扱う上で、機械学習とともに実用性の点から多くの場合に必要になる線形モデル（Linear Model; LM）、混合モデル（Mixed Model; MM）、階層ベイズモデル（Bayesian Hierarchical Model; BHM）について紹介する（1.5.4項参照）。

1.5.2 データと知識ベース

AIが有意な結果をもたらすには、大量のデータ、大量の知識が必要になる。データを生データとすると、ここでの知識とは、データをある程度、整理・加工・抽象化して形式を整え、推論により組み合わせて、有意な結果を導出する源になるものととらえることができる。しかし、データと知識の境界は必ずしも明確であるわけではなく、特に本節で取り上げるビッグデータ時代には、データと知識の中間的なものも増大してきている（例えば自然言語テキストデータ）。

1970年代後半にAIにおける「知識」の重要性が唱えられ、1980年代には知識ベースに基づく多くのエキスパートシステムが作成された（第2次AIブームの時代）。このときの知識の形式は大部分がルール型であり、一部に階層構造を持つフレーム型や論理型、制約型も用いられた。知識の形式はともかくとして、この1980年代の知識の大半は（専門家へのインタビューや資料参照等を介して）人が書き下したものであり、今日のビッグデータ時代とくらべて小規模であり、スケールアップすることが難しいのが大きな問題であった。

人手による知識獲得の問題に対処すべく、1980年代後半から機械学習、データマイニングの研究が活発になり今日に至っている。データマイニングと言っても、ウェブが普及する以前は関係データベースが主なソースで、そこからの規則性発見の研究が多く行われた。

ウェブは情報共有、情報流通のプラットフォームとして登場し、1990年代半ばからのその普及と拡大は、データや知識のグローバルな流通と共有の主要なプラットフォームとして、検索エンジンを伴うことで情報環境に革命的変化をもたらした。2000年代になるとブログやSNS（代表的なものはTwitter、Facebook）なども登場し、いわゆるCGM（Consumer Generated Media）によって各個人のバラエティに富む情報発信も増大した。

知識の点で特筆すべきは、2001年に始まった共同執筆のウェブオンライン百科事典であるWikipediaである。2016年に英語版は約500万項目、日本語版で約100万項目の規模となっている。内容は中立的観点からの記述にすべきとの方針が採られており、情報の質も2005年に百科事典エンサイクロペディア・ブリタニカとの比較で大差はなかったという調査結果も示された。Wikipediaの情報を、コンピュータによる意味把握を容易にするLOD形式にしたのが「DBpedia」である（1.5.3項参照）。

以上のようにして、1990年代半ばから2000年代にかけては、物理世界とは別の情報のサイバー空間が出現し、拡大した時期であった。これに伴い、アクセス可能なパブリックの情報量は拡大の一途をたどっている。

キーワード検索の検索エンジンはウェブ情報空間利用に不可欠のツールだが、雑多な情報の選別など、アクセスには不十分な点も多い。Googleの「Knowledge Graph」は、ウェブ等から抽出したオブジェクト（事物）間の関係を意味ネットワーク形（オブジェクトをノードとし、ノード間の意味的関係を付したエッジで結んだグラフ）で知識化して表したものである。これにより、単なるキーワードでなくオブジェクトの意味的関係も考慮し、曖昧性を回避するような検索を可能にしている。

2012年の発表時点で57億件以上のオブジェクトと、それらオブジェクト間の180億件以上の意味的関係を有している。2016年に発表があったMicrosoftの「Concept Graph」は、テキスト理解に必要な概念をやはりノードとし、ノード間の確率を伴う関係で結ぶグラフとしている。この概念ノード数は540万程であり、テキスト文の常識に照らした確率的解釈に役立つ。1980年代後半から常識の知識ベース化として主に人手に頼って構築が進められたCYC[1]は、およそ50万概念の規模だが、上記Knowledge

---

※1
Cycorp Website <http://www.cyc.com/>

GraphやConcept Graphは、テキスト・ビッグデータからオブジェクトや概念を自動抽出しており、はるかに大規模の知識ベースとなっている（Concept Graphは一般公開される予定となっているが、Knowledge Graphの一般公開は不明である）。

情報のサイバー空間が拡大する一方で、2000年代にはデータ増大に関して別の動向も顕在化した。IoTやCPSなどと呼ばれる、物理空間とサイバー空間の情報を連携、融合させる動きである。これにより物理空間の多種大量のセンサ計測データがサイバー空間へともたらされ、新サービスに使われるようになってきている。気象データ、人の移動や行動データ、人の健康関連センサデータ、各種交通情報データ、物流データ、カメラなどの監視情報データなど、列挙するのは難しいほど多様多種であり、拡大している。

更に企業では、以前にも増して蓄積するデータ量を増している。ウェブ等のオープン・パブリックデータが注目されることが多いが、世界のデータ総量の70～80％程はこのような企業のプライベートデータであるとも言われている。

知識のソースとして、論文文献データも重要である。医療分野を始めとして公表される論文の情報は、人間が見て判断する量を超えており、コンピュータの助けなしでは十分に利用できなくなってきている。その利用も、単なるキーワード検索から、意味的理解に踏み込む利用へと向かっている。

このようにして2000年代半ばにビッグデータ活用の時代になり、今日に至っている。先にも述べたように、AIは大量のデータや知識があって初めて価値を生むことができるのであり、ビッグデータとAIはセットで考える必要がある。

1980年代のルール型知識が主体であった時代の知識の利用法、すなわち推論は、三段論法的な演繹推論が主体であったが、多様なデータや知識のビッグデータ時代はそれらの活用法も多様になってきている。ウェブ上などのテキストデータを活用するには、自然言語理解やテキストマイニングといった手法が主要な役割を果たす。IoTのセンサからの生データの解釈（分類や異常検出など）やデータベース、データの解釈には、従来は統計的手法やパターン認識手法、データマイニング手法が用いられてきたが、ディープラーニングが高性能の成果を達成するようになった2012年以降には、ディープラーニングが注目され、多く用いられるようになった。2016年にニューラル機械翻訳の性能がそれまでのフレーズに基づく統計的機械翻訳の性能を超えるようになり、自然言語処理／理解に関してもDNNは広がる傾向にある。

一方で、ディープラーニングの使用に際して実用面から注意すべき点としては、

（1）高性能は達成できるが学習に従来以上の大量訓練データが必要になること。
（2）判定の過程がブラックボックス的で人間には理解できず、修正が必要になる場合にどこを修正すればよいか分からないことが多い（訓練データを追加して修正しなければならない）

といった点が指摘されている。またディープラーニングは高い成果を得られる可能性もあるが、特にデータ量が必ずしも十分でない場合などは、無暗にディープラーニングを用いるというのでなく、従来からの統計的アナリティクス、機械学習法なども試みて、見通しをつけることが必要になる場合がある（そういった意味で必要になる統計的手法については1.5.4項参照）。

2010年代の代表的なAIシステム及びプラットフォームとしてIBMの「Watson」を例にとり、データ／知識とその活用法を見てみることにする。IBMは、米国の人気クイズTV番組「Jeopardy!」（ジョパディ！）に挑戦するために、4年間の研究の結果としてWatsonを開発して、2011年2月に人間のチャンピオンに優る成績を挙げた。このIBM WatsonはDeep QA（深い質問応答）の範疇のシステムであり、Wikipediaや文献など大量データや知識を有し（100万冊の本に相当する知識を持つとされる）、幅広い分野の質問に対する回答の確信度を伴って出力する。IBMはこの成果をビジネス（IBMではAIで

なく「コグニティブ・コンピューティング」あるいは「Augmented Intelligence」（拡張知能）と称している）につなげるべく、Watsonのブランド名でツールをセットにしたプラットフォームを提供している。

いくつかのツールを挙げると、基礎的なものでは、

・自然言語分類（文脈、意味も解釈した分類）
・テキストデータの検索及びランク付け（機械学習機能を利用した情報検索精度の向上）
・性格分析（人のパーソナリティの分類）
・画像認識
・音声認識と音声合成

などがある。Jeopardy! に挑戦したWatsonが元になっている経緯から、自然言語テキスト処理や理解関連の機能が目立つ。ディープラーニングツールは当初はユーザ向けに提供されていたが、現在は画像認識、音声認識、自然言語処理関連のツール内で使用される形態となっている。やや応用寄りのツールを挙げると

・Watson Knowledge Studio（テキストの分野ごとのカスタムアノテーション機能を提供）
・Watson for Clinical Trail Matching（患者とガン臨床試験適合性を識別）
・Watson for Oncology（多くの情報（データと専門知識）を得てガン専門医治療の判断を助ける）
・Watson Discovery Advisor（医療や法律などで異質なデータソースを調べて洞察を可能にする。学習機能付き）
・Watson Explore（構造化及び非構造化コンテンツを分析し、傾向、パターン、関係等を見出す）
・Watson Engagement Advisor（顧客と対話して学習して知識を蓄積し、質問を聞き、適切な解決策を提示する）

などがある。健康、医療に力を入れているので、この関連の機能が多くなっている。以上はツールと言えるが、以下の二つは大量データや知識を有する独立してサービスを提供するシステムと言える。

・Watson for Drug Discovery（医療論文データMEDLINEの2000万件程の論文を中心に大量の医療学術論文から抽出した成分と効能の関係などの知識を持ち、創薬やライフサイエンス研究をサポートする）
・Watson Genomic Analytics（遺伝子と薬の効き方の関係などゲノムデータに関する膨大な文献データから抽出した知識を持ち、ゲノム情報を利用した治療法（主にガン）の選択をサポートする）

最後の2システムは、データや知識も含めてシステムとして提供しており、今後のAIビジネスの注目すべき方向である。

AIが価値ある結果を生むには大量のデータや知識が不可欠となり、Google、Amazon、Facebook等では自社サービスを通じて得たビッグデータを活用し、ディープラーニングなどのAIを適用することにより次々と新サービスを生み出している。これに比べると、日本企業は所持する、あるいは活用できるビッグデータの点で弱い。政府や自治体がオンライン情報公開を進めるのは良い傾向で、一層進める必要がある。

民間企業では、購買履歴やクレジットなど個人情報、モバイルデバイスから得られる個人の移動や行動情報を所持する企業もあり、自社内では限定的に利用が図られている。しかし、個人情報保護法によ

るプライバシー保護の観点から、それら個人関連情報を外部企業と共有して利用することは一般には難しい。日本企業のビッグデータ不足を補い、Google、Amazon、Facebookに対抗するような多くのAIシステムやサービスを生み出していくには、ビッグデータ共有の枠組みは取り組まなければならない課題である。情報銀行の構想があるが、一つの可能性として期待したい。

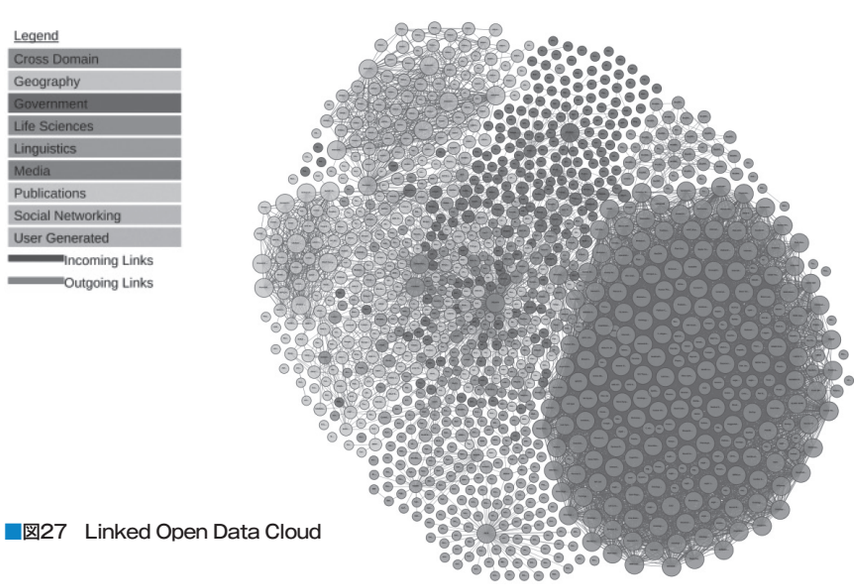
1.5.3 Linked Open Dataとオントロジー

ウェブは、我々の情報流通の仕組みを大きく変化させた。このウェブによる情報流通の革命と同じことが今、データの流通に起きつつある。それが「Linked Data」あるいは「Linked Open Data」（LOD）である。これまでのウェブは主に文書的情報を相互にリンクしてネットワークをつくっている。いわば「文書のウェブ」といえる。これに対して、LODは同様のネットワークをデータの間でつくる。このため、LODは「データのウェブ」と呼ばれる。

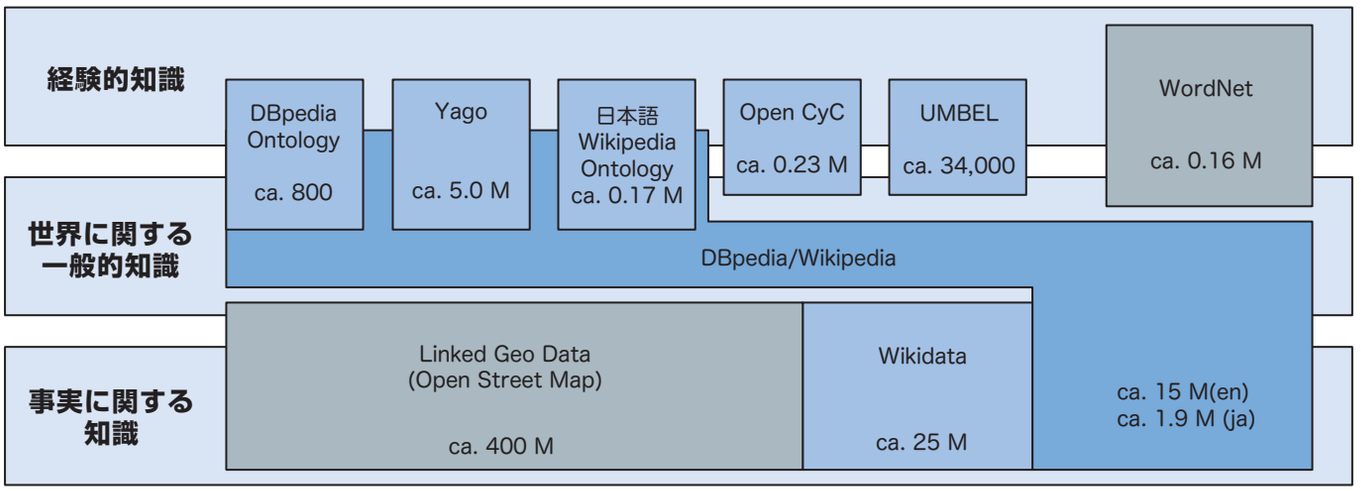
ただし、「文書のウェブ」ではリンクは種類のないものであったのに対して、「データのウェブ」ではデータ間の関係を示すラベルを持つリンクになる。LODはウェブと同じようにグローバルに共有するデータ空間である。すなわち、データが世界のどこのだれが管理しているデータセット（データベース）に含まれているかを意識することなく、アクセスしたり、リンクしたりすることができる。これまであったデータセットとデータセットの間の障壁は存在しない。個々のデータセットはお互いにリンクし合うことで、一つのグローバルなデータセットの一部となるわけである。

LOD技術とは「セマンティックウェブ」研究でつくられた技術をデータの表現に利用したものである。セマンティックウェブとは、ウェブを作ったティム・バーナーズ・リー（Tim Berners-Lee）氏が提唱した、現行のウェブより高度に知識を記述できるウェブをつくるというビジョンである。そのポイントはウェブのグローバルな情報共有空間はそのままに、その上に標準的なメタデータの記法であるRDF（Resource Description Framework）やそのスキーマを記述する言語（RDFS（RDFSchema）やOWL（Web Ontology Language））を用意することで、グローバルに知識を共有する仕組みである。

まず、RDFという言語で全てを書く。RDFは「主語」「述語」「目的語」に相当する3語の組み合わせ（三つ組）で全ての情報を書く言語である。このとき、表現したい事物には個別のURI（Uniform Resource Identifier）を与える。ここでは個々の事物にURIを振ることで、URIが世界中でユニークなIDとして使えることが重要である。更にデータを記述する様式が決まっている場合、その様式をスキーマとして別途定義する。RDFを拡張したRDFSという言語では、データ記述の様式をクラスとプロパティの組み合わせで定義する。更にOWLを用いるとクラスやクラスの階層関係などのオントロジーを表現



■図27　Linked Open Data Cloud



■図28　知識とオープンなデータセットの関係

することができる。OWLは記述論理（Description Logics）に基づいた知識表現言語であるがRDFを用いて記述される。

現在、上記の原則に基づいたLODが多数公開されている。**図27**に示すのは、世界中のLODのネットワークで、これを「LOD Cloud」と呼んでいる[2]。2017年2月時点であり、1000以上のデータセットが含まれている。丸が個別のデータセットを示し、データセットとデータセットを結んでいる線は、データセット内のデータ同士にリンクがあることを示している。

LODはこれまでのウェブと同様に、検索サービスを通じてアクセスして使うこともできるし、リンクを順に辿ってみるというブラウジングでアクセスして使うこともできる。しかし、LODの特徴を活かした使い方は、マッシュアップ・アプリケーションを通じての使い方である。

通常のマッシュアップ・アプリケーションでは、複数のウェブサービスのウェブAPI（Application Programming Interface）を利用してアプリケーションを構築する。LODを使ったマッシュアップ・アプリケーションは通常のマッシュアップ・アプリケーションに比べ、より容易に構築できる。LODはデータのアクセスの仕方が統一されているため、ウェブAPIのように個別に対応を変える必要がない。また、記述形式もRDFで統一されているため、アプリケーションのなかでも統一的に扱うことができる。更に、異なるデータセット間でも、リンクがあればそのまま使うことができるため、データ統合の手間を省くことができる。

LOD技術の利用は、データの公開の場面に限定されているわけではない。データモデルをRDFのみとすることで、多種多様なデータを統合的に処理する「ETL」（Extract Transform Load）[3]の仕組みとして使うこともできる。データの収集、洗練、統合、分析、公開、利用といった一連のプロセスを、RDF処理だけで実現できる。その際、データやスキーマはそのまま外部から取り込むことができ、処理結果のほうもそのまま外部へ提供することもできる。また、様々なRDFに関わるツール、ソフトウェアをその間で使うことができる。

これらのデータセットは統一したフォーマット（RDF）で記述され、また一部はデータ間でリンクが貼られている。この中には事実に関するデータからオントロジーといった知識まで含まれている。これらの知識を大まかに図示すると、**図28**のようになる。

個別の事象、事実に関する汎用的な知識は、Open Street MapのLOD版であるLinked Geo Dataや、

※2
Andrejs Abele et al, “The Linking Open Data cloud diagram.” LOD cloud diagram Website <http://lod-cloud.net/>

※3
多様な情報源から必要な情報を抽出し、適切な形式に変換して統合すること。

Wikidataに大量にある。また個別の事象、事実を抽象化した汎用的な知識はDBpediaやWikidataにある。更に知識を構造化したオントロジーがDBpedia Ontology、YAGO（Yet Another Great Ontology）、日本語Wikipedia Ontology[1]、OpenCyc、UMBEL[4]にある。またWordNetもRDF化されている。重要な点はこれらのデータセットが既に結合されているか容易に結合可能であり、一体化した知識として使うことが期待される点である。実際、IBM Watsonの技術であるDeepQAでは知識としては構造的データと非構造的データ両方が用いられているが、構造化データとしてはRDFで記述されたDBpedia、Freebase、YAGOのデータが用いられている。またCognonto（米国）というスタートアップ企業がWikipedia、Wikidata、GeoNames[5]、OpenCyc、DBpedia、UMBELの6個のデータセットを統合したKBpedia[6]という知識ベースを作っている。

このように、LODはデータと知識をシームレスにつなぎ巨大な知識空間を作る仕組みとして機能している。推論や機械学習とこの巨大な知識空間を組み合わせることが知識に基づくAIの新しい在り方となると思われる。

1.5.4 統計モデル

人が理解できる情報をデータから引出すためには、何らかの理論的裏付けを持った処理をデータに施す必要がある。これは何もビッグデータに限ったことではなく、100サンプル程度の比較的小規模なデータの解析においても同じである。ただし、データの規模が大きくなると顕現する問題もある。ここでは、古典的な線形モデルを出発点とし、ビッグデータ解析に付随する問題を解決し得る統計モデルを紹介する。

1.5.4.1 線形モデル

説明変数を含む項の線形結合で目的変数を記述するモデルで、基本となるモデルは目的変数（y）も説明変数（x）も1次元の場合、すなわち

$$y = b_0 + b_1 x$$

である。これは単回帰（single regression）に等しい。このとき、各項の係数（b_0、b_1）をパラメータと呼び、測定したデータから推定する。なお、目的変数が連続値（連続尺度）の場合は回帰（regression）だが、離散値（名義尺度）の場合は分類（classification）と名称が変わる。

実際にはデータの説明変数が1次元であることは少なく、2次元以上にわたることが多い。この場合、説明変数の次元数（n）に応じて線形結合する項数を増やし

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$$

とモデルを拡張する。このように、説明変数の次元が2以上の場合を重回帰（multiple regression）と呼び、各項の係数パラメータを偏回帰係数（partial regression coefficient）という。

偏回帰係数では目的変数へのある説明変数の影響度を計ることはできない。変数の尺度が変われば、

※4
Umbel Website <http://umbel.org/>

※5
GeoNames Website <http://www.geonames.org/>

※6
Cognonto Website <http://cognonto.com/>

偏回帰係数の尺度も変わるからである。目的変数と説明変数をそれぞれ平均0、分散1に標準化した場合の偏回帰係数を標準偏回帰係数（standardised partial regression coefficient）と呼び、統計モデルのそれぞれの説明変数の影響度を比較することが可能となる。

モデルから得た目的変数の推定値と実際のサンプルの目的変数との誤差を残差（residual）というが、上記モデルの残差が正規分布に従う場合は一般線形モデル（general linear model）となる。この残差の分布の制限を取り外し、任意の分布と置くと、上記モデルは一般化線形モデル（Generalized Linear Model; GLM）へ拡張される。なお、係数パラメータの推定には、LMでは最小二乗法を用いるが、一般線形モデルやGLMでは最尤推定法を用いる。

1.5.4.2 混合モデル

ここまで紹介してきた統計モデルは、目的変数（y）の中央値（median）を推定するものである。しかし、データによっては例えば測定誤差が大きく、線形モデルではうまく推定できない、あるいは、推定できても推定値のばらつきが大きくなってモデルの信頼性を担保できない場合がある。そこで、説明変数（$x_1 \sim x_n$）に新たな概念を導入し、このばらつきをうまく吸収するために考案された統計モデルが混合モデルである。

混合モデルでは、説明変数に固定効果（fixed effects）と変量効果（random effects）が混合していると考える。説明変数の固定効果は目的変数の中央値に影響し、説明変数の変量効果は目的変数の分散、すなわちばらつきに影響する。この変量効果は測定誤差に限らず、データを得た場所や時間、個体によって説明変数のばらつきに差がある場合も含まれる。例えば、心血管疾患の罹患（Cardio Vascular Disease; CVD）を目的変数、健康診断項目の血圧（Blood Pressure; BP）、体重（weight; W）と身長（height; H）を説明変数とそれぞれ置いた重回帰を考える。最も単純な統計モデルは

$$CVD = b_0 + b_1 BP + b_2 W + b_3 H$$

という線形モデルである。ところが、血圧には季節変動があり、冬に高くなることが分かっている。したがって、このデータの説明変数には固定効果（真の血圧、BP）と変量効果（季節変動、e）が混在しており、混合モデルを採用することが望ましい。

$$CVD = b_0 + b_1 BP + b_2 W + b_3 H + e$$

混合モデルをGLMに当てはめたものが一般化線形混合モデル（Generalized Linear Mixed Model; GLMM）である。変量効果は多くの場合は観測できない潜在変数として現れるので、係数パラメータの最尤推定[7]にはEMアルゴリズム（expectation-maximization algorithm）が利用されることが多い。

1.5.4.3 階層ベイズモデル

GLMMはサンプルの個体差や場所・時間による効果の違いをモデル化できる優れた方法であるが、更に自由に統計モデルを記述したい場合もある。例えば、複数の変量効果をモデルに組み込む場合、あるいは、データの構造が入れ子関係になっている場合などである。具体例を挙げると、ある会社の従業員の健康診断データを解析する際、従業員は事業所や職種で分類されると同時に、更に年齢層や性別で

※7
確率分布に含まれるパラメータを、与えられたデータの分布を生み出す確率が最も高くなるように決定する手法。

分類される。このように階層的なデータ構造をモデルに反映し、事業所間での相似や差異、事業所の違いを問わない性差などを検討したい場合である。

階層ベイズモデルは、上記のような複雑なデータ構造を自在に記載し、入れ子関係となっているグループごとに異なる変量効果を設定することができる。また、ベイズモデルは固定効果の係数パラメータの生成も確率過程に従うと仮定し、係数パラメータの分布を推定する。つまり、前式のb_0～b_3はあるパラメータを持った確率分布（例えば平均muと分散vの正規分布）、すなわち確率変数として推定される。係数パラメータのパラメータをハイパーパラメータと呼ぶ。こうしたパラメータの推定にはマルコフ連鎖モンテカルロ法（Markov Chain Monte Carlo methods; MCMC）を利用することが多い。

1.5.4.4 スパースモデリング

ビッグデータではサンプル数の増大とともに、説明変数の次元数も増大する傾向にある。説明変数の次元数が増えると線形モデルの項数が増え、可能なモデルの個数は直ちに膨大な数になる。多数のモデルから最適なモデルを探索するとき、（1）多重性、（2）モデル探索の実行可能性の、二つの問題が浮かび上がる。

多重性の問題は、統計モデルに対して検定を繰り返すと、実際には意味のない統計モデルを有意であると誤認してしまう危険性が増すということである。また、膨大な数の統計モデルの全てを試すことは計算時間の爆発的増加を招き、現実的な時間内での実行が困難となる。これが実行可能性の問題である。

これらの問題は、何らかの方法で説明変数の次元数を削減すれば、一定の解決が見込める。説明変数の多くは目的変数の予測に寄与せず、少数の説明変数で組み立てた統計モデルで十分である。そういう寄与しない説明変数の係数パラメータを0と置くことでデータを疎行列化し、データの次元を圧縮する技術がスパースモデリング（sparse modeling）である。コアになる技法はL1正則化と呼ばれ、統計モデルに罰則項を追加して行う推定である。罰則項のハイパーパラメータは正則化の効果を決定し、交差検証（cross validation）で決める。

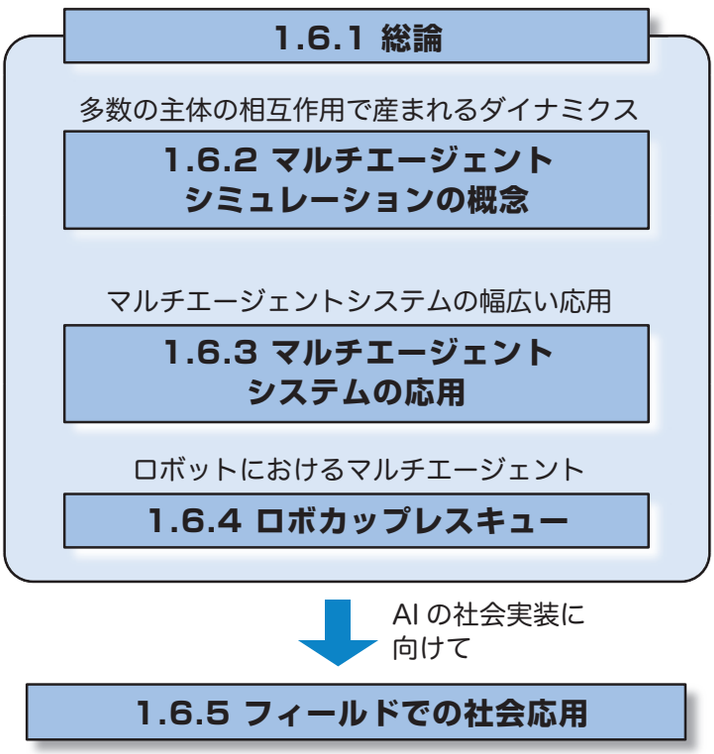
線形モデルではL1正則化はlasso（least absolute shrinkage and selection operator）と呼ばれ、寄与の低い説明変数の係数パラメータは0となる。いわゆるモデル選択／特徴量選択を行ったこととlassoの使用は同義である。スパースモデリングは特にビッグデータの解析に大きく貢献し、技法が発表されてから20年以上経った現在でも活発に研究が続けられている。

ここで紹介した統計モデルのより詳細な解説や実装は、文献[2][3][4]を参照されたい。

参考文献
[1] 玉川奨ほか「日本語 Wikipedia からの大規模オントロジー学習」『人工知能学会論文誌』vol. 25 No. 5, pp.623-636.
[2] 久保拓弥『データ解析のための統計モデリング入門 一般化線形モデル・階層ベイズモデル・MCMC』岩波書店.
[3] 岩波データサイエンス刊行委員会『岩波データサイエンス Vol. 1 ベイズ推論とMCMCのフリーソフト』岩波書店.
[4] 岩波データサイエンス刊行委員会『岩波データサイエンス Vol. 5 スパースモデリングと多変量データ解析』岩波書店.

1.6 社会とコミュニティ

1.6.1 総論



■図29　本節の構成

本節では人工知能（AI）の社会応用について述べる。実際の応用までを目指した研究においては、AIと他の技術（例えば情報通信技術）などが一体となることが多く、ITの社会応用と呼べる側面が強い。逆に、現在IoT（Internet of Things）やサイバーフィジカルシステム（Cyber Physical System; CPS）の名で呼ばれている分野にも、AI的要素は多い。ここではそれらの中から、特にAI的要素の強いテーマを選んだ。

一般論として技術は基礎研究と応用の間を往復しながら成長する。応用によって基礎技術の不足が判明し、その研究に戻ることもあれば、新しい技術の芽がそれまで不可能だった応用を可能にすることもある。基礎と応用はスパイラル進化をする。

AIに関していえば、一般にAIの夏といわれる時期には基礎研究が盛んになり、その応用がなされる頃には世間的な注目を集めることが少なく、AIの冬といわれる。この理由の一つはAIが応用可能なところまで枯れた技術となる頃には、それはAIとは呼ばれないということに起因する。社会応用がなされる頃には単なる技術であって、AIではないと認識されることが多い。文字認識などが好例であろう。あまり正確に文字認識できない頃にはAIの研究として盛んに新しい手法が試される。しかし郵便番号や宛名の自動読み取りが可能になる頃にはその適用範囲の方に目が向いてしまい、それを可能としている技術自身は語られなくなる。

「マルチエージェントシミュレーション」（Multi-Agent Simulation; MAS）（1.6.2項参照）は実用時にもその技術が認識され得る数少ないものかもしれない。従来は、経済学では人間集団を理想的な行動

を取るものとして理論化してきた。それ以外の分野でも集団を統計的に扱うことしかできなかった。

これに対して、MASでは個々のエージェントの判断を個別にシミュレートできるため、実社会に近いモデルが構築できる。現在のところ、MASは研究ツールとしては経済学に始まる様々な分野でその威力を発揮しているが、主としてシステムの分析あるいはデザイン段階でも用いられることが多く、実用システムに取り入れられた例は少ない。

交通システムや人流シミュレーション（1.6.3項参照）は、実用化されてもシステムの裏でMASが走るという数少ない例かもしれない。前者は公共交通網をITとAI（特にMAS）で管理・運営することにより、現在より飛躍的に効率の良い運行が可能となり、利用者の利便性も増し、自治体の住民サービス向上の一環として住民（特に高齢者）のモビリティ確保の強力な手段となる。

米国では、国防予算の一部がAI研究に投入されている（1.9.2項参照）。これは日本における国費のAI研究投入に比べて1桁から2桁多いと言われている。日本では国防のための研究には反対意見も多い。ならば防災はどうであろうか？　こちらは大義名分があるにもかかわらず、やはり国費投入は限定的である。しかし、研究者の間では防災研究への意欲は大きく、「国際救助隊」設立をスローガンとしてロボカップレスキュー（1.8.3項参照）の枠組みが立ち上がった。

1990年代後半からAIの冬の時代が始まる。その一方でユビキタスコンピューティングや環境知能といった分野が始まり、現在のIoTやCPSへとつながっている。様々な分野名で呼ばれているが、これらは全てコンピュータ上のモデルと実世界を様々なセンサやアクチュエータでつなぐことにより、現実世界での人間活動を支援することを目的としている。この頃に開始されたものに「フィールド情報学」[1]がある（1.6.5項）。様々な現場で情報技術の研究と適用を行う分野である。IT（あるいはAI）をフィールドで活用するときに、最近急激に力をつけているディープラーニングは、環境認識を担う「AIの眼」として大いに期待されているから、この分野は今後違った方向への発展が期待される。様々な社会情報システムから得られる膨大なビッグデータを実時間で解析、学習しながらAIシステムが稼働するようになれば、実用化してもAIの名が残るようになりそうである。

1.6.2 マルチエージェントシミュレーションの概念

マルチエージェント社会シミュレーション（Multi-Agent Social Simulation; MASS）は、小さなAIプログラム（エージェントと呼ぶ）を人間とみなし、それが多数集まって相互作用することで生じる現象をもって、人間の社会現象を模倣しようという試みである。人間社会と同じく、個々のエージェントは各々異なる能力や性質、目的を持って独立に意思決定するが、道で出逢えばお互いに避けるなど、互いに影響を受けながら判断するものとされる。シミュレーションの主眼は社会の模倣にあるため、エージェントは非常に賢いものである必要はなく、社会現象を再現する程度に知的であればよい。

シミュレーションの対象となる社会現象としては、人流、物流、交通、情報流通、金融、経済活動など多岐にわたる。またエージェントも、人そのものの場合から、アルゴリズムトレードにおけるプログラムなどの人の代わりとして行動するもの、更には会社組織なども仮想的な「人」エージェントとして扱う場合もある。

MASS技術の出口は、制度設計あるいはサービス設計の支援、なかでも、新規の制度やサービスがシステム全体に与える影響を分析することである。この分析についても、個々の状況における個別の評価よりは、多様な状況下でのシステム全体のマクロな状態について、その挙動の分類や不具合状態に陥る条件の洗い出しが重要なテーマとなる。

例えば、人流シミュレーションを使った避難誘導方法の評価を考えると、ある特定条件における、ある手順での避難の効率を評価したり、あるいはその条件での最適な避難方法を調べたりすることは、原

理的には可能である。しかし、災害という、そもそも状況を詳細に予見することが困難な対象を相手にしていることを考えると、ある特定条件のみにこだわるのではなく、幅広い様々な状況について、安定して機能する誘導方法を見つけ出すこと、あるいは、懸案となっている誘導方法が機能しなくなる条件はどういう場合かを調べることが、一つの大きな応用方法となる。

MASSで問題となるのは、シミュレーションとしてのモデルの精度と結果の信頼度（Validation and Verification）である。エージェントは人の振る舞いをモデル化・模倣するものであるが、人の振る舞いについてはまだ確実なモデルといえるものはほとんどなく、物理法則のように第一原理で全てを精度よく予見できる段階にはない。また、人間は、自分の行動の結果が自分にとって不都合であることを知れば行動を変えてしまうかもしれない、という自律性がある。つまり、シミュレーションを行うこと自体が系に影響を与えてしまうというジレンマを包含しているのが、社会シミュレーションの特徴である。このため、社会シミュレーションを台風の進路を推定するのと同じようなもの、と考えることは現状では危険である。

一方で、シミュレーションの結果をマクロにとらえ、系全体としての性質を分析するツールとみなすことで、MASは非常に有用なツールとなり得る。例えば自動車交通の中でどのような条件で渋滞が生じ、どのように解消していくのか、というマクロな視点で分析すれば、どの条件が渋滞の鍵となっており、どのような方法で軽減するかの方策などについて議論することができる。

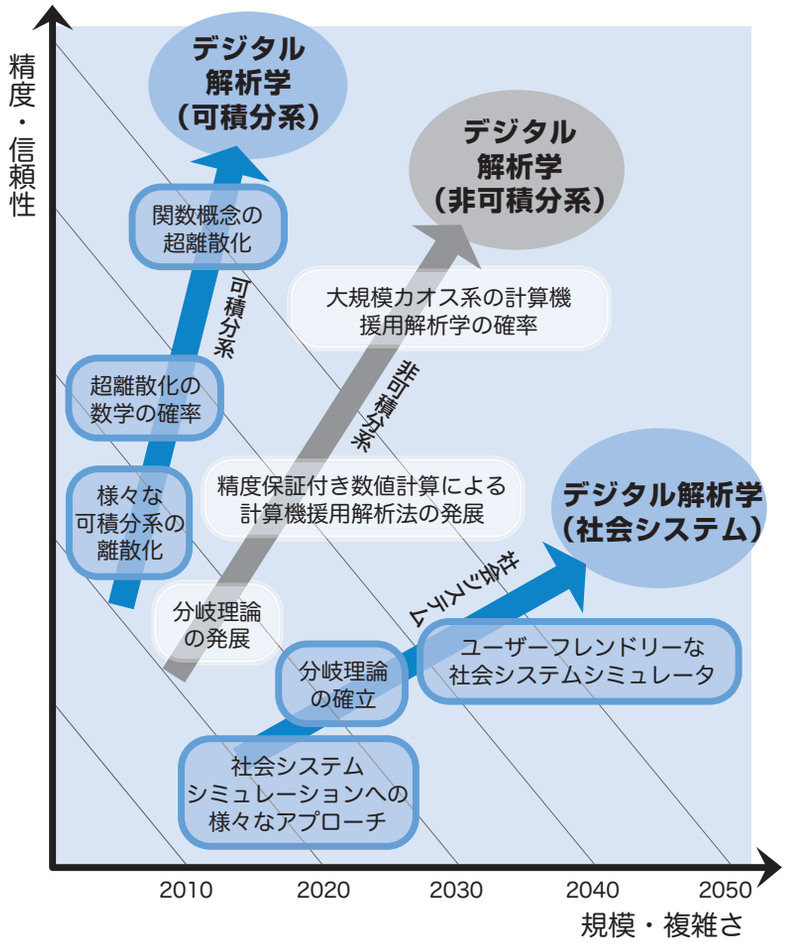
これを可能にするためには、多数の異なる条件で対象の系がどのように振る舞うかを網羅的にシミュレーションし、その結果を解析していく方法が大事になってくる。2009年の横断型基幹科学技術研究団体連合の分野横断型科学技術アカデミック・ロードマップ報告書（第4章）[1]でも述べられているように、今後のMASS研究では、複雑系としての分岐理論の確立に向けた取組を進めていく必要がある（**図30**）。このためには、応用分野や現象を広げていくほか、モデルの精密化につながるセンシングデータとの関連付け、データ同化手法、及び網羅的シミュレーションの知能化・効率化に関する研究及び環境作りが重要となってくる。

MASの応用分野は幅広いが、特に相性の良い分野の一つとして、経済システムのモデル化がある。なぜなら、現実の経済現象は、様々な思惑を持った個人の行動が互いに影響を及ぼしながら集積した結果現れた現象だからである。まさに、MASが解明しようとしている創発的現象の主な例である。また、経済現象では、各個人の行動基準が比較的明確である。金融市場ならば取引によって利益を出すことであり、購買行動ならば自分にとって価値の高いものをできるだけコストをかけずに購入することである。そのため、エージェントの行動ルールも素直に設計しやすい。

更に、近年の情報通信技術の発達のおかげで、より広範囲で迅速に、詳細で時間分解能力の高い社会経済データが収集できるようになってきた。大量かつ大規模な実データを用いて、経済現象の背後にある個人行動の分析を行い、その結果をエージェントの行動ルールに反映させることによって、エージェントの行動をより現実的にすることができる。

MASの経済への応用は、金融市場、マーケティング、オークション、計算組織理論、製品や技術の普及、電力市場、環境経済など、多くの分野にわたっている。各分野において、実際の経済現象のメカニズムの解明や従来の経済理論の検証、そして現実の経済現象でのルールの設計などに用いられている。

※1
横断型基幹科学技術研究団体連合「分野横断型科学技術アカデミック･ロードマップ報告書」経済産業省ウェブサイト
<http://www.meti.go.jp/policy/economy/gijutsu_kakushin/kenkyu_kaihatu/20fy-pj/oudan2.pdf>



■図30　複雑系シミュレーション技術のロードマップイメージ[2]

1.6.3 マルチエージェントシステムの応用

1.6.3.1 新しい交通システム

MASの社会応用の一つとして、より効率的な都市公共交通のデザインがある。移動の要求を出すユーザやユーザを輸送する車両などをエージェントとして実現。現実の市街地の地理情報の上で、公共交通の時刻表、移動需要の発生、天候などに関する情報を加味することで、現況再現性の高いシミュレーションが可能となってきた。市街地の地理情報に関する標準フォーマットOSM（OpenStreetMap）、オープンでフリーのエージェントベースの交通シミュレータSUMO（Simulation of Urban MObility）[3]、MATSim（Multi-Agent Transport Simulation）[4]などが普及している。

近年、ユーザに移動というサービスを提供する「MaaS」（Mobility as a Service）と呼ばれる概念が提唱されている。移動させることをサービス提供の一種とみなすことで、何かほかのサービスとの共創やハードウェア、ソフトウェアとのパッケージ化を促すことができる。MaaSの導入を進めているフィンランドのヘルシンキでは、タクシー、バス、鉄道、飛行機などの輸送機関が互いに協力、情報交換す

※2
横断型基幹科学技術研究団体連合「分野横断型科学技術アカデミック・ロードマップ報告書」経済産業省ウェブサイト
<http://www.meti.go.jp/policy/economy/gijutsu_kakushin/kenkyu_kaihatu/20fy-pj/oudan2.pdf >

※3
“SUMO – Simulation of Urban MObility” DLR - Institute of Transportation Systems Website <http://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/>

※4
MATSim Website <http://www.matsim.org/>

るための情報システムと仕組み作りを行っている。MaaSを社会実装できれば、ユーザは、そのような移動サービスを単一のインターフェースで利用できるようになる。

MaaSの概念を押し進めると、移動サービスのクラウド化という概念に至る[2]。タクシー、バスなどではサービスとそれを提供するハードウェアが分かちがたく結び付いているが、それらを分離して需要に応じて再結合させることができれば、必要な時に必要なだけの移動サービスを生み出すことができる。一般に、クラウドという概念が、CPUや記憶装置を仮想化してネットワーク経由で計算や保存というサービスとして利用できるようにしたシステムを意味するように、移動サービスのクラウド化は、タクシーやバスといった移動手段を仮想化し、需要に応じて同じハードウェアで異なる移動サービスを提供するシステムを意味する。

「Uber」や「Lyft」、「コンビニクル」[3]などが新しい交通システムとして話題に上ることが多い。これらは、移動の要求を持っているユーザがタクシーを呼んだり、自分で選択した公共交通機関を利用したりする行為において、従来のワークフローのデジタル化に留まっている。移動サービスのクラウド化、全体最適という共創のレベルまでは至らず、デジタル化の恩恵は限定的である。

移動サービスのクラウド化は、より効率的な都市公共交通システムを実現する可能性があるが、そのためには、全体最適な車両の運行管理、ユーザからの需要の処理、需要予測に基づく最適化などを解決せねばならない。これらの課題解決には、システム設計→MAS→結果分析→システム改良・設計→…のサイクルが欠かせない。また、現在タクシー会社の熟練オペレータが担当している配車、アクシデント発生時のバス運行緊急対応、個々のタクシー運転手の勘に任されている流し待機エリアなどは、機械学習が得意とする対象でもある。

そういった、乗客のデマンド（呼び出し予約）に応じて、乗降時刻あるいは乗降場所を柔軟に調整しながら走行する公共交通の形態を「デマンド交通」と呼ぶ。バスだけでなく相乗り方式のタクシーや、両者を混在させたデマンド交通の活用例も増えている。デマンド交通は、自由度に応じて大きく以下のような運行方式にカテゴリ分けすることができる[4]。

① 迂回路方式：路線バスとして基幹ルートを運行しながら、事前のデマンド受付により迂回ルートを適宜経由、定められた停留所で乗降する方式。
② フレックスルート方式：あらかじめ停留所は定めておくが、停留所候補をつなぐルートはデマンドに応じて決める方式。
③ ドアトゥドア方式：停留所や路線を定めずに、乗客が指定した任意の乗降場所を運行する方式。
④ フルデマンド方式：ドアトゥドア、ダイヤフリー、リアルタイム予約（運行中車両へのデマンドも可能）のデマンド運行が可能。すなわち「いつでも、どこでも、すぐに」を実現する方式。

従来は、自宅や目的地で乗降できる自由度の高さから、③をフルデマンドと呼ぶ場合もあった。だが、現実には「毎時運行」などのダイヤ固定や、「1時間前までの予約」など運行上の制約条件が多く、また運行ルートをオペレータや運転手が経験知で決定していることも多いので、ここではドアトゥドアと呼んだ。更に上記③と④については、予約システムと配車システムがコンピュータで自動制御されているかどうかの違いもある。現状では多くが人力に依存しているために、明確な分類定義がないが、これについても定義しておく。

(i) 予約・配車いずれのシステム導入もなし
(ii) 予約支援システムを導入
(iii) 配車運行自動制御システムを導入

⑷ フル自動制御（予約＋配車運行）システム

運行方式とは別の軸として、運行目的・運行区域による区分も考えられる。

（a） 特定運行型（Special Transport Services; STS）
過疎地域や高齢化地域、交通空白地域等の、特定区域・特定目的の運行を目的とするもの。
（b） 広範運行型（General Transport Services; GTS）
中心地域や特定地域という区分を超えてより広範なエリア、広範な目的に対してデマンド交通を導入しようとするもの。

特定運行型は欧米など海外でも使われている定義だが、広範運行型はここで新しく導入した定義である。従来のデマンド交通はほとんどが特定運行型であり、そもそも「広範に運行する」という考え方が存在しなかった。両者の違いは運行エリアの大小ではなく、むしろ都市交通全体の中での位置付けが末端的なのか（a）、中枢的なのか（b）ととらえるほうが適切であろう。

次に国内外の主なデマンド交通の先進事例を取り上げ、実用化の状況を概観する。

まず公共交通へのデマンド交通の導入で、最も成功しているといわれるスウェーデンでは、政府が全土で高齢者・障害者向けの乗合タクシーをコミュニティ単位で運行してきた。ストックホルムやヨーテボリなどの中核都市では、これをフレックスルート方式のデマンドバスへ発展させ、運行地域を急拡大させてきた。

2007年時点でストックホルムでは人口185万人に対して年間利用回数450万回、ヨーテボリでは人口56万人に対して170万回といずれも高利用率だが、位置付けはあくまで中心市街地に対する末端部、誰でも利用可能だが主には高齢者居住者が多い地域での運行である。ヨーテボリの「flexlinjen」（flex line）は、2005年の8路線から2010年には20路線にまで拡大している。予約受付は電話で、オペレータがデマンドのあった停留所をインプットし、運行ルートを運転手に知らせるシステムである。降りる場所はデマンドではなく中心市街地への結節点となる停留所が想定されている。

フィンランドの首都ヘルシンキでは、市街地中心部でフレックスルート方式のデマンドバス「Kutsuplus」の導入が試験的に始まっている。ユーザの現在地と目的地に最も近い停留所とマップの提示、バス乗継を含めた複数ルートの提案、ルートや所要時間に応じた価格設定などが、携帯端末から利用可能になっている。フルデマンド方式ではないが、都市中枢部のコミュニティバスをフル自動制御でスマート化することが目指されている。

米国では、シリコンバレーの「RidePal」、ボストンの「Bridj」などが、乗合のデマンドバスサービスに進出しているほか、日本でもタクシー呼出システムの進出が話題となっているUberが、マイカー相乗りサービスも展開している。米国で民間主導のサービスが見られるのは、欧州と異なり政府主導の公共交通の役割が限定されていることが背景にある。そのため、公共の基幹交通と、民間主導のデマンドサービスとの統合が遅れていることが課題といえる。

日本においてもデマンド交通の導入は全国各地で進んでいる[5]。国土交通省の地域交通ネットワーク政策[5]において、基幹ルートは定時路線運行で、結節点から枝分かれする末端部分をデマンド交通とするモデルが描かれているように、実用化事例のほとんどが特定運行型で、迂回路又はフレックスルート方式のコミュニティバスが多勢を占めるが、ドアトゥドア方式やフルデマンド方式のミニバスや乗合

※5
「地域公共交通活性化事例集」国土交通省 地域公共交通支援センターウェブサイト <http://koutsu-shien-center.jp/jirei/>

タクシーの運行も増えてきている。

国内の数少ない事例の一つとして、ドアトゥドア方式でかつ広範運行型を志向しているのが、北海道ニセコ町「にこっとBUS」、岩手県一戸市「いちのへ いくべ号」、岡山県総社市「雪舟くん」[6]である(**表3**)。ニセコ町と総社市は、路線バスの存続が厳しくなり、市町全域をデマンド交通でカバーする広範運行に移行した例である。一戸市の場合は既存の路線バスとの共存がなされている。

■表3　国内のデマンド交通：ドアトゥドア方式×広範運行型の事例

名称	運行地域	運送法上の分類	経営主体	運行事業者	スケジュール	予約方法	料金(おとな)	車両
ニセコ町 デマンドバス 「にこっと BUS」	北海道ニセコ町 ほぼ全域	区域運行	ニセコ町	ニセコバス	ダイヤなし (午前 8 時〜午後 7 時/毎日運行)	一週間前から概 ね 45 分前まで に電話予約	1 乗車 200 円	10 人乗り 2 台
「いちのへ いくべ号」	岩手県一戸市 全域	乗合タクシー	LLP 一戸市 デマンド交通	バス 1 社、 タクシー 3 社 (左記出資元)	ダイヤなし (平日 7：30〜 16：30)	2 日前〜1 時間 前までに予約	区域制： 300〜700 円	ジャンボなど 4 台
「雪舟くん」	岡山県総社市 全域	区域運行	総社市	バス 2 社、 タクシー 5 社	平日 8 時台〜 16 時台まで 往復計 16 便	1 時間程度前 までに予約	1 乗車 300 円	10 人乗り 4 台、 8 人乗り 5 台

ニセコ町では、365日ダイヤなしでドアトゥドアのデマンドに対応するが、要事前予約、運行台数2台と、広範運行とはいえ小規模な運用である。しかしながら、配車運行システムを導入して、どの利用者にも大きな不便を与えることのない経路をシステムで選定し運行している。

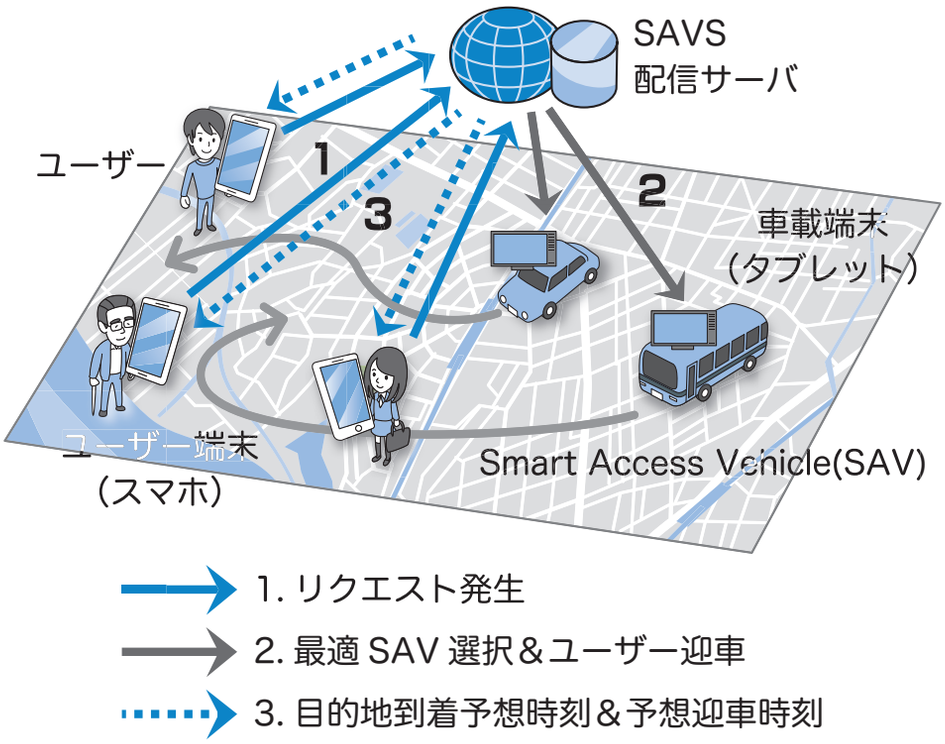
総社市の場合、乗降場所は自由に指定できるものの、出発地域（周辺部）から到着地域（中心市街地）への方向性をもち、毎時1本運行と本数も決まっているため、実態はフレックスルートのドアトゥドア版という形式になっている。また、土日運行はなく、平日の高齢者用途を主目的としているため、特定運行の拡張版ともいえる。予約も配車運行もオペレータの手作業と経験知で行われている。

一戸市も、平日運行ながらダイヤなしでドアトゥドアのデマンドに対応しているが、要事前予約で台数も4台と少ない。広範運行ではあるものの、町内のほとんどを占める中山間地域から中心部へのアクセス改善を目的として、既存の路線バスを補完する運用となっている。

現在開発が進められているSAV（Smart Access Vehicle）Systemは、④フルデマンド×（b）広範運行型×（iv）フル自動制御を兼ね備えた「デマンド応答型公共交通」（Demand Responsive Transport）の一種であり、固定経路を持たずユーザの呼び出しに応じて即時に配車される[7]。このとき、先に乗車している、しようとしている乗客が乗合いを許容し、その乗客の目的地到着希望時刻が守れるのであれば、乗合いが発生する。つまり、従来のタクシーとバス両方のサービスを兼ね備えたような交通システムである。

ユーザがSAVを呼び出す時の動作を**図31**に示す。まずステップ（1）にて、ユーザはスマートフォン上のアプリケーションを通じて、SAVS配車サーバにリクエストを上げる。このとき、移動サービスのクラウド化による全体最適化を実現するため、ユーザは事前に移動サービスに関する全ての条件をサーバに通知しなければならない。ユーザのリクエストに対し、配車サーバが最適なSAVを選択し、ステップ（2）にて当該SAVにユーザをピックアップする指令を送る。SAV運転手は、SAVに積んだ車載端末（タブレット）上のアプリケーションでその指令を受信し、ユーザを迎えに移動する。同時に、配車サーバはユーザに対し、目的地到着の予想時刻、迎えに行くSAVの現在位置を通知する。

SAVSには次の三つの特徴がある。一つめは、ユーザからのリクエスト応答処理に、予約ベースでなく即時実行ベースを採用したことである。これにより、SAVSは、例えばバスのような固定路線、固定ダイヤの運行形態をエミュレートできる。つまり、多数のユーザが毎回同時刻に同じ場所間を移動するデマンドを発出し、それをSAVSが即時に処理すると考えれば、バスと等価な運行形態が実現される。同様に、従来タクシーをエミュレートすることも、デマンドバスをエミュレートすることもできる。



■図31　SAVSの呼び出し・配車の基本動作

二つ目は、過疎地域ではなく、少なくとも中規模都市より大きいエリアとユーザ数を想定していることである。シミュレーション結果より、ある規模を超えると現状のタクシーやバスによる乗客輸送システムの効率を上回る領域があることが明らかになっている。ユーザは移動サービスに関する条件を全て事前にSAVSに通知し、計算機がそれを考慮して完全自動で最適な配車を決定するので、原理的にSAVSを展開するサイズが大きくなればなるほど全体最適化の余地が増える。

三つ目は、乗客輸送を制御するパラメータ、例えば移動経路、出発時刻、乗合の可否などを動的かつ連続値として設定できることである。これにより、ユーザの要望、その場の状況や制約に応じて、柔軟かつきめ細かいサービスレベルを実現することができる。一般の交通システムに同じくSAVSでもコストと利便性の間にはトレードオフがあるが、SAV車両ごとに乗客ごとに輸送に関するパラメータを動的に変更してサービスレベルを調整して、そのトレードオフに対処する。また、そのパラメータ変更を受けて全体最適化の制御を変更することになる。

SAVSの本格的な研究開発は2012年から始まり、2017年まで函館市や東京お台場にて実証実験を4回実施した。最大規模で同時に30台のSAV、200人超の乗客からの乗車リクエストを完全自動で遅滞なく処理し、移動サービスを提供することに成功している。

1.6.3.2 人流シミュレーション

人流シミュレーションは、群衆が町中や施設の中を歩きまわる状況を再現しようとするものであり、MASSでは特に、群衆を構成する人の多様性や個々人の経験の評価に着目してシミュレーションするものである。人の多様性としては、その場所に関する知識や認識（土地鑑）、歩く速さや分岐点で経路を選ぶ際に重視する要素（目的地までの距離・混雑度・道の雰囲気など）、あるいは出発地・目的地や経由地の違い等が考えられる。経験としては、旅行時間のほか、混雑に巻き込まれていた時間や、目的地・経由地への到着時間などがある。例えば人流シミュレーションの一つ、避難シミュレーションでは、最終的な避難場所やそこへの経路情報、あるいは周辺の他エージェントの挙動や混雑度などが情報として

与えられる。これらにより進行する方向を決め、近傍の障害物や他エージェントとの相対距離により実際の歩行速度を求め、移動する。

また、音声や情報機器による情報提供を、行動決定のための情報として与える場合もある。多くの場合、各エージェントは独立して行動するよう設定されるが、シミュレーションによっては、家族などを単位とした集団で行動したり、他のエージェントに追従したりして動くといった性格付けを行う場合もある。

人流で、エージェントが移動する空間については、連続的な2次元平面、平面をグリッド上など小領域に分けたセル空間、廊下や道路を中心とした点と線のネットワークとして扱うモデル、更には、そのネットワーク上の各交差点における待ち行列のみに着目したモデル、という抽象化レベルがあり得る。前者ほどエージェントの細かい挙動を記述でき、スクランブル交差点や広場のある空間での避難現象を正確に扱うことができるが、多くのパラメータ設定と計算量が必要となる。一方、後者は大規模な避難のような移動を比較的簡便に扱うことができるが、人々が入り乱れて移動するような現象を正確には扱えないという問題がある。このため、対象とする避難の規模や求められるシミュレーション速度に応じて、適切な抽象化を選ぶ必要がある。

シミュレーションによる人流解析は、災害時の避難など、まれな状況や新規の状況での人流の分析に強みがある。近年ではモバイル空間統計など、人流に関係するビッグデータの利用が容易になりつつある。これらのビッグデータは平常時に頻繁に繰り返される現象を分析・モデル化する上では有用であるが、災害など滅多に起きない状況や、新しい設備や道路の導入など新規の状況においてはビッグデータそのものでは対処できない。また、災害状況などをリアルに再現した実証実験などは危険性の面から難しいため、実データを実験で揃えることも困難である。

一方、シミュレーションでは、その状況を自由に設定できるため、新規の状況やまれな状況設定に容易に対応できる。更に、そのような分析は災害対策立案やイベントでの全体管理設計において、当事者に全体的な把握を促し、判断する際などにおける有用な情報を提供できる可能性が高い。

AIの研究から見た場合の人流シミュレーションの焦点は、エージェントの行動、特に目標地点や経路の選択である。単純な人流シミュレーションでは、各エージェントが目的地（避難場所など）やそこに至る最短経路を既知とし、エージェントは寄り道せずに目的地に向かう。しかし実際の人の挙動は、目的地を特に持たなかったり、経路が分からなかったりする場合があり、経路選択において揺らぎ・混乱が生じる。また、混雑状況や誘導アナウンスなど、ほかから得られる情報を元に行動をするケースも考えられる。人流シミュレーションでは、各エージェントにそのための知的な判断ルーチンを個別にもたせることができるため、混乱の波及や情報提供の効果などを詳細に分析することが可能となる。

1.6.3.3 金融分野への応用

「人工市場」とは、金融市場のエージェントモデルである。1990年代後半に、複雑系研究で有名な米国のサンタフェ研究所が、人工市場に学習と創発という観点を初めて導入し、バブルの発生や予測ルールの複雑化を分析した。2000年代に入り、人工市場は、より詳細で現実的になった。国内の人工市場研究プロジェクトの一つである「U-Martプロジェクト」は、多様な取引プログラムが市場サーバに接続する人工市場である。これにより、手数料率や値幅制限などの市場制度のテストを行った。「AGE-DASI TOF」[6]は、現実の経済記事を基にしたデータを人工市場に入力し、現実のある時期のバブル発生・崩壊のメカニズムを解明したり、市場介入政策の決定を支援したりするシステムを構築した。

※6
多数の仮想的なディーラーが参加する外国為替市場を模擬した人工市場。A Genetic-algorithmic Double Auction Simulation in Tokyo Foreign exchange market.

金融市場における制度設計や投資手法の変化は、その市場の安定性に大きな影響を与えるが、新たな制度の導入による市場インパクトを事前に知ることは困難である。そこで、人工市場を構築し、制度の効果検証を行う取組が行われている。応用事例は、次の3とおりである。

(a) 板寄せ（バッチオークション）やザラ場（連続ダブルオークション）などの異なる価格決定方式（市場に出された注文をマッチングし価格を決定する方式）を採用した人工市場により、それぞれの市場効率性や価格変動の性質を比較したもの。
(b) 空売り制限やサーキットブレーカー、取引税などの市場安定化のための様々な規制の効果を人工市場で検証したもの。
(c) NASDAQや東京証券取引所に人工市場を適用して、新しい制度（価格変動幅の最小単位の切り下げ）の投資家へのインパクトを検証したもの。

1.6.3.4 マーケティング分野への応用

近年の情報通信技術の発展により、マスメディア、ソーシャルメディア、口コミなど、様々な情報源から膨大な情報を消費者が参照できるようになった。それに伴って、各消費者の選択ルールや消費者間の相互作用がより複雑で多様になり、既存の消費者行動モデルでは説明できない現象が現れている。

そこで、消費者の購買行動とコミュニケーション行動をモデル化したMASにより、消費者行動のミクロな変化と商品売上などのマクロな動態との関係を分析する応用事例が数多く発表された。例えば、市場の動きがエージェントの意思決定過程と製品特性、及び社会ネットワーク構造（社会ネットワーク中のハブのサイズ）の三つの要素間の相互作用に依存することを示した研究がある。他の研究では、消費者の情報チャネルの増加により、各消費者の選択が特定の財に集中する傾向を強める可能性があることが示された。このように、消費者間の新たな相互作用がマーケットに与える影響が、エージェントシミュレーションにより様々な視点から明らかにされている。

消費者側からの視点だけでなく、生産・流通側からの視点によるエージェントシミュレーションの応用事例も数多くある。特に、近年の消費者行動の多様化により、より詳細で高度な生産・在庫管理の手法が提案されている。これらの新たな生産・在庫管理システムを実装したシミュレーションが、複数の市場環境の下でのシステムの有効性（利益、コスト、機会損失）を評価することに使用されている。

1.6.3.5 電力・エネルギー経済分野への応用

近年の世界的な動きとして、電力を含むエネルギー経済分野において様々な規制を緩和し、市場競争の導入による効率化を目指すという自由化の流れが進んでいる。更に、大規模災害の発生により電力網の安定性に大きな関心が集まり、スマートグリッドや分散電源などの、従来よりも柔軟で複雑な新しい電力供給システムが着目されている。

この分野でのMASの主な応用事例は、電力卸売市場と送配電システムに関する新制度を、安定性と効率性の両方の観点から評価するものである。電力市場に関しては、国内外の多くの研究でエージェントシミュレーションが使われ、既に米国ではエージェントベースの大規模な電力市場が構築され、電力市場制度の研究で利用されている。

送配電システムについても、実際の電力需要データや発電データを用いて電力消費エージェントの挙動を決定しスマートグリッドシステムの効率性を評価した研究や、分散した小規模な電力市場により電力価格と配電ネットワークを創発的に構築する新たな電力流通システムの有効性を分析している研究などがある。

1.6.3.6 製品や技術の普及過程分析への応用

既存のものと大きく異なる革新的な製品や技術が、どのような過程で世の中に広まっていくのか、また時には普及せずに終わってしまうのかは、経済分野での大きな関心の一つである。イノベーションの普及は、消費者間の相互作用というマーケティングの要素、企業間の競争関係という組織戦略の要素など、複数の要素が関係する創発的現象である。そのため、消費者や企業を対象としたエージェントモデルを利用して、イノベーション普及過程の解明を目指す研究が増加している。

例えば、規模の小さな初期市場から主要市場に発展するまでの溝（キャズム）の分析、消費者間の複雑ネットワーク構造が与える競合サービス普及への影響分析、企業間競争と社会ネットワークの相互作用の観点からの技術普及の分析などがある。

1.6.3.7 環境経済への応用

社会における環境意識の高まりとともに、資源を効率的に利用する循環型社会のための新制度を設計する必要性が高まっている。この分野でのエージェントシミュレーションの応用は、家電リサイクル法など生産者及び消費者にリサイクルや回収に関する何らかの義務を課す制度の有効性を評価するものや、環境に配慮した製品の市場への普及過程を分析したものなど、制度の社会的評価や消費者行動の分析を対象にしたものである。

1.6.3.8 マルチエージェントシミュレーションの経済への応用の新しい方向性

エージェントベースの経済シミュレーションのベースとなる個人の行動データの収集については、観察技術の発達によって、より詳細で精緻なデータが、日常の社会的場面でリアルタイムに取得できるようになってきた。今後は、前述の経済現象に関する表面的な行動データだけではなく、脳科学（brain science）の発達により、経済行動の背後にある個人の認知機構や思考過程、行動決定ルールに関する詳細な分析技術の発達も期待される。

特に、2000年代半ばから機能的MRI（fMRI）を始めとする新しい脳観測技術を、経済的選択時の脳活動の分析に用いる神経経済学（ニューロエコノミクス）の発達が目覚ましい。それと同時に、マクロな経済状態についても、より広範囲で高頻度な観測データがリアルタイムに利用可能となった。これらのデータを基にして、複合的な経済シミュレーションモデルをリアルタイムに構築・更新することが可能となり、様々な現実の経済現象の中で、エージェントシミュレーションが現実世界に浸透していく方向に進むと考えられる。

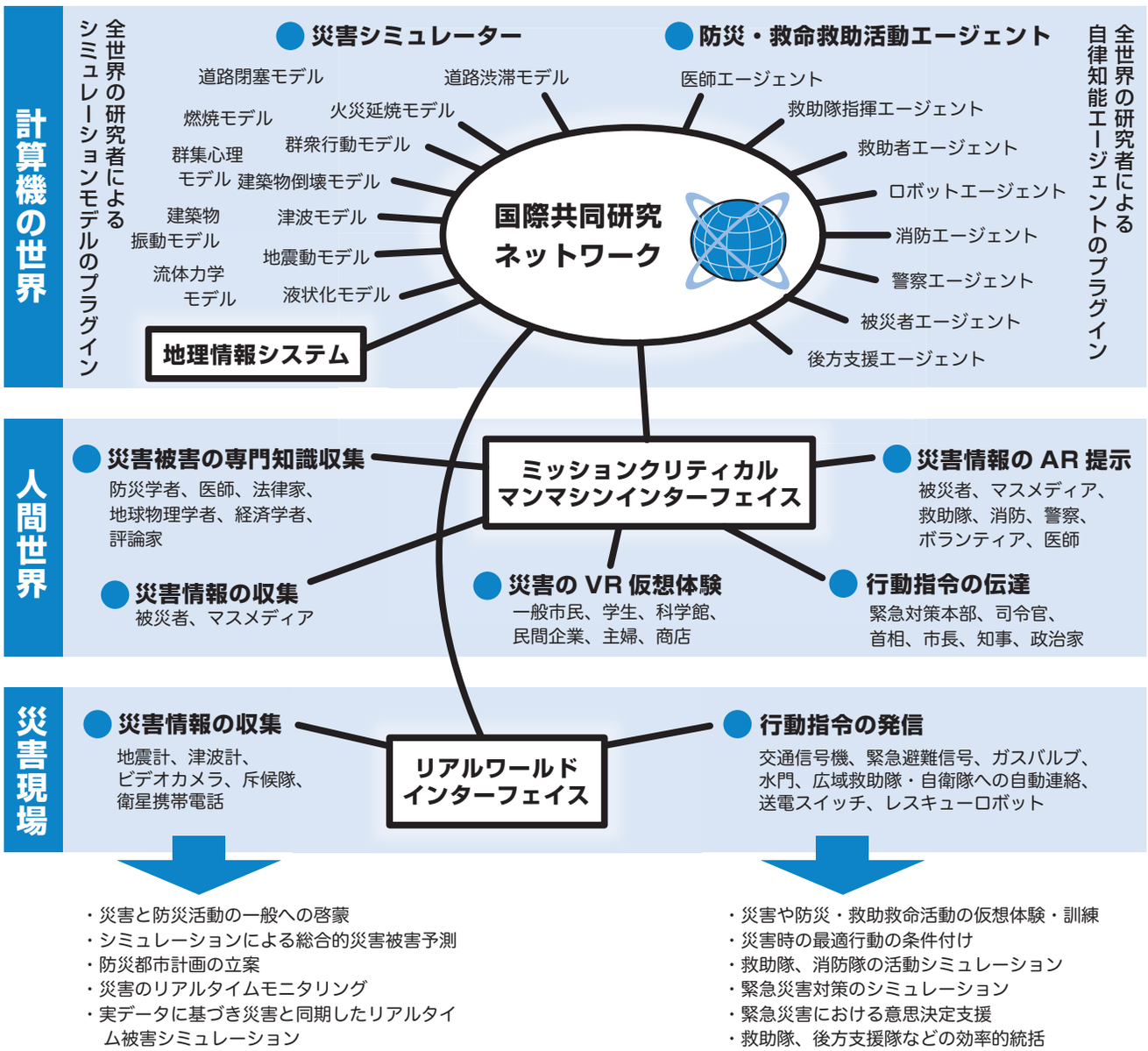
1.6.4 ロボカップレスキュー

1.6.4.1 ロボカップレスキュー構想

ロボカップ（RoboCup、1.8.3項参照）の中に災害救助部門を作ることを目的として、1998年に共同研究コミュニティが立ち上がり、AIやロボットを活用した未来の防災システムの在り方についての議論が行われた。その成果は「ロボカップレスキュー」（RoboCup Rescue）構想として**図32**のようにまとめられ[8] [9]、その一部を競技会として開催すべく、ロボカップレスキューが開始された[7]。

災害事象の分散シミュレータによって作られるサイバー災害空間の中で、災害緊急対応活動を行うエージェントが活動することによって、自然現象と社会現象を合わせた総合的な災害シミュレーション

※7
RoboCup Website <http://www.robocup.org/>



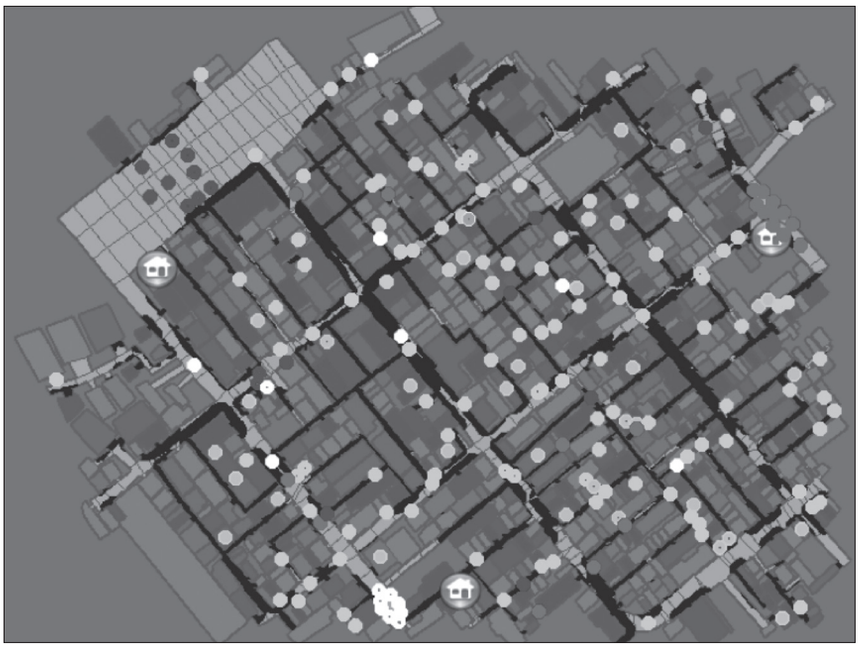
■図32　ロボカップレスキュー構想

を行う。サイバー空間はリアルワールドインターフェースを通じて社会に配備された災害状況センシングシステムからの情報をリアルタイムに収集し、シミュレータの状態量に反映させることによって、仮想空間と現実空間の同期を可能にする。

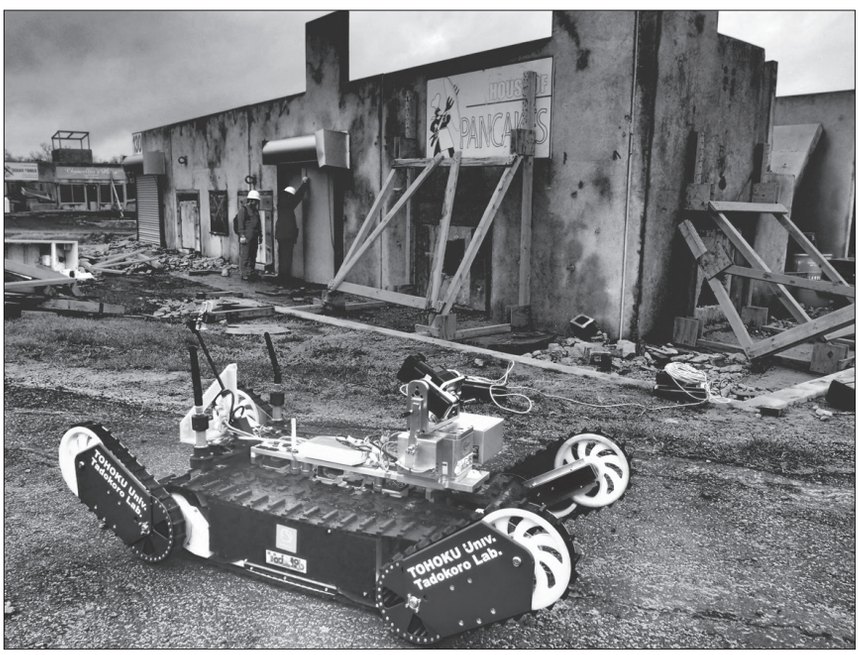
また、行動指令を発信することによって、各種インフラやロボットなどがリアルタイムに能動的な情報収集や災害被害を抑えるための活動を行う。更にミッションクリティカルマンマシンインターフェースによって、災害対応本部の意思決定を支援する。それによって、災害被害軽減のための最適活動を合理的かつ実時間で行うことが可能になる。

1.6.4.2 ロボカップレスキューシミュレーションリーグ

ロボカップレスキューシミュレーションリーグは、阪神淡路大震災における神戸市長田区の被災状況をモデルとし、市街地規模での救助活動を競うシミュレーション競技会であり、ロボカップレスキュー構想に従って2001年に開始された。設定可能なシナリオ（地図及び災害の発生状況）に基づいて被災状況を再現し、消防隊、救急隊、土木工事隊及び各司令所からなる多数の自律型エージェントプログラ



■図33　ロボカップレスキューシミュレーション



■図34　Quince

ムによって災害救助活動を行い、人命や地域の被害などの減災効果で救助エージェント方針の優劣を決定する（**図33**）。

家屋・道路の被災状況や被災による被害者など多岐にわたる減災効果を指標にし、マルチエージェントアルゴリズム（不完全状況下での資源割当、チームワーク、情報共有等）の競争的発展促進と、その成果に基づく社会貢献を目指した。その間、災害下の救助活動シミュレーションの問題点である対象とする社会問題の複雑さと人の行動規範の定式化、シミュレーション結果の再現性・検証について、シミュレーション対象を明確にするルール及び規模拡大に対するマルチエージェントシステムアーキテクチャの提案、エージェント開発フレームワークの導入による対応が行われてきた。

ロボカップレスキューによるシミュレーションは対象分野の重要性により、大都市大震災軽減化特別プロジェクト（2002～2006年）の震災総合シミュレーションシステム[10]や、英国のALADDIN

（Autonomous Learning Agents for Decentralized Data and Information Networks）プロジェクト[10]、米国でHomeland Security Awardを獲得したARMOR、IRIS、GURADSプロジェクト等に影響を及ぼした。

1.6.4.3 ロボカップレスキューロボットリーグ

ロボカップレスキューロボットリーグは、直接目視できない遠隔から競技用ロボットにより災害現場を模したフィールド内にある被災者を探索し、その数と状態、それらの場所を示した地図作成の精度と速さを競う競技である。災害空間としては不整地、急な階段、複雑な形状の障害物などから構成される迷路が、被災者として穴の開いた箱に隠れた人形が音源や熱源とともに設置され、ロボットに搭載したセンサによって情報を収集する。ロボットは自己位置推定と地図作成を同時に行いながら（Simultaneous Localization and Mapping; SLAM）、被災者の探索を進める[11]。被災者の捜索能力の競争に加え、完全自律型ロボットは自律行動や地図生成の能力を、遠隔操縦ロボットは不整地移動性能やヒューマンインターフェースの能力を競う。

競技が開始された2001年にはカメラを搭載したラジコンタンクのレベルに過ぎなかったロボットが、ステップフィールドと呼ばれる高さの異なる角材を組み合わせた不整地や急な階段、障害物の間を自由自在に遠隔あるいは自律で行動できるようになり、自動的に被災者を発見して地図を生成し、入り組んだ場所にある物体を遠隔操作できるようになるなど、被災者捜索の技術レベルを高めることに貢献した。

本競技会は災害ロボットの研究分野の創生と活性化に貢献し、それまでは研究テーマとして取り上げられることが皆無であったものが、現在では多くの研究者がロボティクスの活用分野の一つとして取り上げ、その高度化に取り組むこととなった。その成果は、福島第一原発の原子力建屋に国産第一号機として投入された「Quince」の開発につながるなど、実際の災害対応にも一定の成果を上げている[11]。

その競技ルールは、米国国立標準技術研究所（NIST）によって、災害ロボットの性能評価法としてASTM（American Society for Testing and Materials）標準となり、災害対応ロボットの調達や隊員の操縦訓練に活用されている[8]。また、DARPA Robotics Challenge、euRAthlon、World Robot Summit（ロボットオリンピック）など、世界中で災害対応を目的とした競技会が開催される基盤となった。

1.6.4.4 残された課題

ロボカップレスキュー構想は、当時リアルタイム防災と呼ばれた自動災害情報収集システムの次世代の形として注目されたが、それを実現するための技術レベルや社会インフラが未熟であり、実際の防災組織や手順との整合性がとれないことから、現実的とはいえなかった。

提案から20年近くを経て、今、構想実現のための技術基盤が整いつつある。地理情報の整備が進み、様々なものが情報化され、災害予測精度は飛躍的に向上し、大規模シミュレーションを可能にする枠組みが整ってきた。無線ネットワーク、IoT、ロボットの発展により、大規模に情報を収集し、行動を指令するための基盤が整備されてきた。あらゆる仕事はネットワーク化され、携帯電話や仮想現実（Virtual Reality; VR）などのヒューマンインターフェースの普及も著しい。また、自律ロボットが活動できる環境が飛躍的に拡大して生活空間における実用性が確保され、災害のような不定環境においても限定的に作業を実施できる見込みが高まってきた。

ここでの最大の課題は、ほかのAIの問題と同様、シミュレーションのサイバー空間やロボットが持つ

※8
"Standard Test Methods for Response Robots." NIST Website
<http://RobotTestMethods.nist.gov>

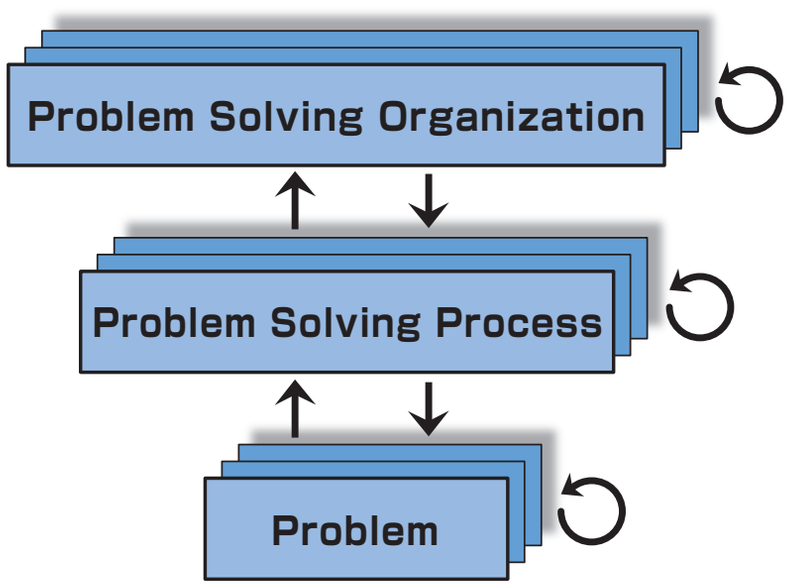
世界モデルと、リアルワールドとの間に存在する、大きな溝を埋めることである。機械学習等の手法によって、実データの解釈の実用性が徐々に高まってきている。人間の組織活動も、ルーチンワークについては少なくとも情報化が進んできている。そのため、10年のスパンで、ある限定された範囲においてロボカップレスキュー構想が実現し、安全で安心して暮らせる社会の実現に貢献すると予測される。

1.6.5 フィールドでの社会応用

我が国は少子高齢化を始めとする課題先進国と呼ばれているが、それらの課題はひとときに解決できるものではなく、継続的な問題発見と問題解決を必要とするフィールドととらえたほうがよい[1]。フィールドは「分析的、工学的アプローチが困難で、統制できず、多様なものが共存並立し、予測できない偶発的な出来事が生起し、常に関与することが求められる場（片井修）」である。この定義は、工学的に解決可能な問題を切り出すことが不可能だと言っているわけではないが、様々なステークホルダーが関わる問題の時空間的な広がりが、重層的な問題空間を構成し、仮に切り出したとしても予期せぬ事象によって次々と変容し、いつまでも手を抜くことができないと指摘している。機能分化された社会のなかでも、最も機能分化されたコミュニティに属する研究者や技術者にとって、フィールドは手を焼く存在である。

同様の指摘は、デザインの分野でも古くからいわれている。「Wicked problem」というのは、不完全で変化する要求に対してデザインすることの困難さを示す表現である。問題間の相互依存性が原因となり、一つの問題を解こうとするとその過程で、また別の問題が現れる有様を表している。AIの分野では、フレーム問題が議論されてきた。エージェントの行為の影響を、おそらくは影響を受けないであろう事象を省いて表現するにはどうすればよいかという問題である。先のフィールドの定義は、フィールドにおける問題がWicked problemであり、それに解を与えることはフレーム問題を経験的に解く困難な試みであることを示唆している。

課題を抱えるフィールドで、問題を継続的に切り出し、解決するために、問題、問題解決プロセス、問題解決組織を以下のようにとらえる（**図35**）。



■図35　問題、問題解決プロセス、問題解決組織

- 問題は、フィールドにおける課題から、解くことを前提として切り出される。
- 問題解決プロセスは、字義どおりには、問題に対して解を与えるプロセスである。しかし、問題は設定された時点で解くことが前提とされるので、実際には、問題解決プロセスは解ける問題を切り出すプロセスであることが多い。そのため、問題発見という言葉が使われることもあるが、解くことを前提とした問題の発見は問題解決と表裏一体である。
- 問題解決組織は、問題解決のためのステークホルダーのネットワークである。言い換えれば、解くことを前提とした問題を継続的に切り出すための組織である。

フィールドにおいて切り出された問題は相互に依存していることが多い。そうであれば、問題を解くプロセスも相互に依存せざるを得ない。その結果、問題解決の主体である問題解決組織も相互に依存することになる。加えて、問題、問題解決プロセス、問題解決組織の依存関係は時間の経過にしたがって変容する。フィールドにおける、問題空間の重層的な関係は、問題を固定し永久的な解を与えることを難しくする。そうであれば、問題の変容に追随できる柔軟な問題解決プロセスをデザインしなければならない。すなわち、眼前の問題を解くことを最終の目標とするのではなく、継続的に問題を解くプロセスのデザインを目標とすべきである。更に、そうしたプロセスを実行する、変化に柔軟な、おそらくはネットワーク型の問題解決組織を維持するべきだろう。

具体的な事例として、途上国支援における多言語コミュニケーションの例を挙げる。途上国支援などで専門家が現地就業者に技術情報を伝える場合は、専門家が現地に赴き、対面で現地就業者に伝える、という方法が一般的であった。しかしながら現地就業者が非識字者であることも多く、伝えた専門知識を将来に向けてそのコミュニティに蓄積させたり、近隣に拡散させたりすることが困難であった。一方、近年では世界的に教育制度が向上していることもあり、途上国の郊外でも児童の就学率が向上し識字率も高くなってきている。そこで、重要で専門的な知識を、ICTを使って国内外の専門家からオンラインで児童を介して非識字の保護者に届ける「YMC」（Youth Mediated Communication）と呼ばれる新しい途上国支援モデルが提案された。

YMCモデルを用いて、ベトナムのメコンデルタ地帯の農家を対象に、日本人専門家による知識提供の支援が行われた。高い精度の翻訳を実現するために、農業分野の辞書や用例対訳が作成され機械翻訳に適用された。しかし、それ以上に努力を要したのは問題解決組織の形成である。ベトナムの農業農村開発省（MARD）、ベトナム・ヴィンロン（Vinh Long）省の農業農村開発局（DARD）、ベトナム国家大学、特定非営利活動法人パンゲア、京都大学の言語グリッドチーム、東京大学や三重大学の農業支援チームがネットワーク型の組織を形成し、2011年から2014年の間にヴィンロン省のTra On地区とBinh Minh地区において4回の実証実験（延べ16か月）が実施された。各回の実証実験には15～30世帯の農家・児童が参加した。日本人農業専門家がベトナム児童を経由して、ベトナムの農民に農業知識を伝えた。要するにこのプロジェクトの中で、機械翻訳技術が、その継続的な利用を実質化する問題解決組織の形成を伴って用いられた。

フィールドと研究者の協働は痛みを伴うこともある。フィールドにしばしば見られる新規技術に対するアレルギーと、研究者の技術的楽観論（情報技術は10年で性能が100倍に向上する）は相容れない。情報システムの人為的複雑さによる説明限界や、情報ビジネスの予測不能性（無料メールがいつの間にかビジネスになる等）が、研究者とフィールドとのコミュニケーションを難しいものにする。フィールドは、研究者にとって活動しやすい場とは限らない。しかしながら、今後、イノベーションの多くはフィールドから生じる。総合的研究領域である情報学においては、研究者のフィールドへのアプローチは必須である。

参考文献

[1] 京都大学フィールド情報学研究会『フィールド情報学入門 ―自然観察、社会参加、イノベーションのための情報学―』共立出版.

[2] 中島秀之ほか「バスとタクシーを融合した新しい公共交通サービスの概念とシステムの実装」『土木学会論文集D3(土木計画学)』vol.71 No.5, pp.I_875-I_888.

[3] 大和裕幸ほか「オンデマンドバス:公共サービスに於けるイノベーション－オペレーションズ・リサーチ」『経営の科学』vol.51 No.9, pp.579-586.

[4] 鈴木文彦『デマンド交通とタクシー活用』地域科学研究会.

[5] 中島秀之ほか「新しい交通サービス実践への道程」『サービス学会第3回国内大会講演論文集』

[6] 田柳恵美子ほか「デマンド応答型公共交通サービスの現状と展望」『人工知能学会全国大会』

[7] 中島秀之ほか「Smart Access Vehicle System：フルデマンド型公共交通配車システムの実装と評価」『情報処理学会論文誌』vol.57 No.4, pp.1290-1302.

[8] 田所諭・北野宏明(監修)・ロボカップ日本委員会(編)「RoboCup-Rescue技術委員会・The RoboCup Federation」『ロボカップレスキュー ―緊急大規模災害救助への挑戦―』共立出版.

[9] Hiroaki Kitano and Satoshi Tadokoro, "RoboCup-Rescue: A grand challenge for multi-agent and intelligent systems," *AI Magazine*, vol.22 Issue.1, pp.39-52.

[10] 後藤洋三「震災総合シミュレーションシステムの開発」『大都市大震災軽減化特別プロジェクト総括成果報告書』文部科学省, p.83.

[11] Keiji Nagatani et al., "Emergency response to the nuclear accident at the Fukushima Daiichi Nuclear Power Plants using mobile rescue robots," *Journal of Field Robotics*, vol.30 Issue.1, pp.44-63.

1.7 計算インフラを構成するハードウェア

1.7.1 総論

インターネット上のトラフィックは増大し続けており、2020年には40ゼタバイトに到達すると考えられている。我が国のデータ流通量も、2005年の約1.6エクサバイトから2014年の約14.5エクサバイトまで、約9.3倍に伸びており[1]、今後も増大トレンドは変わらないと考えられる。人工知能（AI）の対象領域も今後拡大するものと考えられることから、AIの計算速度、低消費電力、低メモリ等への要求は今後も増大し続けるのは間違いない。そのような要求から、AIの研究開発と実用化を支える基盤として、AIの計算等に係るインフラストラクチャの整備やハードウェアの研究開発が進められている。

AIの研究開発や実用化に際して、HPC（High Performance Computing）技術ベースのクラウド的な共有計算環境を整備していくことが、我が国における新たなプレイヤーの参入ハードルを下げ、AIの今後の普及を支える基盤の一つとなる。AIの学習時には、現状では一般に豊富な計算資源が必要となるが、一部の大企業を除いて大規模な計算環境を整えるのはコストの面から難しいため、共有可能な環境を整備することが望ましい。

CPU、GPU（Graphics Processing Unit）、FPGA（Field Programmable Gate Array）、あるいはディープラーニング用の専用チップなどの計算デバイスに関しても、増大し続ける性能に対する要求に応えるべく研究開発が必要となる。更に、計算を実行するための環境だけでなく、データやモデルの共有など、計算周辺の役割も含めた環境整備も合わせて検討を行うことが必要となると考えられる。

AI、特に機械学習の利用のためには、①データから学習によってモデルを作成するフェーズ、②学習後のモデルを用いて新たなデータに対し推論を行うフェーズがある。学習時には、大量のデータをメモリにロードし、反復しながら精度を高めていく計算が必要であるため、計算性能が重要視される。一方、推論時には、個々の入力データに対し、比較的少数回の演算を行えば結果が得られることから、計算性能に対する要求は学習時ほど高くはないが、データの入出力や格納、転送に関し高い能力が求められる。

したがって、AIの学習、特に大規模な計算環境が必要となるディープラーニングでは、例え大企業であっても個別の会社で計算環境を取得・維持・管理することは一般に困難であり、国が実施する機械学習向けのHPCベースの大規模計算環境の整備や、それらの技術を移転した民間によるクラウドの形態での計算リソースの提供の必要性が今後増していくと思われる。

学習時の計算デバイスに対する性能や性能／電力比の向上への要求は、推論時に比べれば必ずしも高くなく、性能優先で利用されるため、現状ではGPUが広く使用されている。ただし、消費電力低減の要求やメモリ容量に対する要求を無視して良いわけではないため、今後FPGAの利用や、専用チップの開発を推進していくことも求められる。一方でGPU等の汎用チップ自身も機械学習向けの性能や省電力性能を増しており、今後競争の激化が予想される。

しかしながら、現状では学習の速度を上げるのは、応用数理及びHPC由来のアルゴリズム面の進化と、同様のハードウェアでの並列処理に強く依存している。アルゴリズムとしての並列性は、ネットワークのフィードフォワードなどのプロパゲーション（1.7.2項参照）を計算する際の行列の積などのアルゴリズムの低レベルな並列性、プロパゲーション時にネットワークを分割し、分散並列計算するモデル並

※1
「平成27年版　情報通信白書」第2部ICTが拓く未来社会、「我が国におけるビッグデータ流通量の推計」総務省ウェブサイト
<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/html/nc254310.html>

列性、バッチ学習時にバッチ用のデータを分割し、学習勾配を後に統合するデータ並列性、更に一種類のデータセットに対し様々なモデルを組み合わせて多数決を取るアンサンブル学習や、そのほかの探索により学習のパラメータやネットワークの種々の構成を決定するハイパーパラメータチューニングのような上位の並列性がある。

これを実際のマシンに実装する際には、物理的にはチップ内の並列と、複数のチップを組み合わせた計算ノード内での並列、更にはノード間の学習の速度を上げるにはスーパーコンピュータ由来の相互結合網で高速に結合されたノード間の並列を活用する必要がある。つまり数十万～数千万の階層的な並列性を、それを実行可能なマシンに実装するわけであるが、それはまさにHPCであり、スーパーコンピュータの劇的な進化に伴い、ディープラーニングに代表される機械学習が革命的に進化し、新たなAIの隆盛をもたらした由来である。つまり、AIの進化には、単にチップだけでなく、スーパーコンピュータ由来の「システム」や「インフラストラクチャ」の設計・開発・構築・配備が大変重要となる。

推論時には、計算性能への要求よりもむしろ性能や性能／電力比への要求が高まる。特に、実用化の際には、推論の比重が高まる。そのため、クラウド・フォグ・エッジという階層構造の中で、エッジ側に近づくほど、低消費電力、低メモリ、短レイテンシ[2]に対する要求が高まり、機械学習に特化したFPGAや専用チップの開発が有利になる可能性がある。

しかし、ディープラーニングのソフトウェア側の研究は日進月歩であり、ハードウェア構成を設計する際には、積和演算など研究が進んでも不変な部分と、ニューラルネットワークの組み方など現在も様々なモデルが模索されている部分の仕分けが重要であると考えられる。このほか、計算への負荷を軽くするための技術として、計算に用いるビット数の削減の研究がある。ニューラルネットワークの計算には必ずしも通常の数値計算で使用される単精度（32ビット）の精度は必要ないことが分かってきており、8ビットや、極端な場合は1ビットでの計算とすることも研究されている。このような計算は、FPGAで比較的容易に実現可能であるため、FPGAの試験的な利用が短期的には増大するものと考えられる。しかしながら推論では有効なものの、学習時には単精度ぐらいの精度は一般的には要求されるとの研究もあり、このあたりはまだ決着していない。

GPUや高速結合網に代表されるHPCの加速技術が、機械学習の加速にも適用可能であるゆえ、スーパーコンピュータによる加速がまずは期待される。しかしながら、「京」に代表される現状の我が国のスーパーコンピュータ群は、歴史的、技術的な理由などにより機械学習やそれを必要とするビッグデータ処理に最適化されておらず、かつ既存のシミュレーション主体のワークロードで既に容量が手一杯であり、AIに提供できる実際の資源量は乏しい。その中で、東京工業大学学術国際情報センター（GSIC）の「TSUBAME」（Tokyo tech Supercomputer and Ubiquitously Accessible Mass storage Environment）シリーズはAI、ビッグデータに処するハード、ソフトの最適化設計を行っており、2017年8月に稼働する最新のTSUBAME3.0[3]により、従来のスーパーコンピュータの汎用ワークロードと共有ではあるものの、そのディープラーニングの総合性能は単体では47.2ペタフロップス、既設のTSUBAME2.5等との合算では65.8ペタフロップスと、我が国トップの学習能力となる。

産業技術総合研究所人工知能研究センター（AIRC）、理化学研究所革新知能統合研究センター（AIP）、情報通信研究機構（NICT）ユニバーサルコミュニケーション研究所（UCRI）など、各省庁関連のAI関係の研究センターや、主だった企業には数百テラフロップス（TFlops[4]）～数ペタフロップス（PFlops）

※2
通信における遅延時間。

※3
「東工大のスパコンTSUBAME3.0が今夏稼働開始―半精度演算性能47.2ペタフロップス、人工知能分野における需要急増へ対応―」東京工業大学ウェブサイト <http://www.titech.ac.jp/news/2017/037500.html>

の小・中規模の機械学習専用のマシンが近年導入されつつある。

また、国が本格的に実施する計算のための大規模環境の整備計画の例としては、産業技術総合研究所（AIST）の「AI Bridging Cloud Infrastructure（ABCI）」がある[5]。ABCIでは、TSUBAME由来のHPC技術を駆使し、130ペタフロップスの機械学習のための専用クラウド計算環境を構築し、アカデミアと民間企業のオープンイノベーションを推進することが目指されている。ABCIでは、計算環境の提供だけでなく、データや学習後モデルの共有なども含めた総合的なプラットフォームとする計画である。ABCIは専用のデータセンターを構築し、2018年3月頃に稼働が予定されている。

なお、民間によるクラウドの形態での計算リソースの提供に関しては、海外の情報系の企業が先行していたが、国内でもさくらインターネットなどによる「高火力コンピューティング」などのサービスが開始されている。

ディープラーニングの計算には、現状ではGPUを利用することが一般的であるが、中期的には、FPGAの利用、専用チップの開発等を含め、学習時や推論時の要求機能・性能の違い、エッジ、フォグやクラウドの利用環境の違いに合わせて多様な構成のデバイスが求められるようになるであろう。

例えば、Googleは推論に限定されるが、TPU（Tensor Processing Unit）と呼ばれるディープラーニング専用のチップを開発し、既に利用していることを明らかにしている。データセットが大きくなるにつれて、推論と学習に要求される資源量の差が本質的に広る。そして、推論はエッジ側での組み込みコンピューティングでの利用が中心的になる一方、学習はクラウド側でHPC由来の大規模なコンピュータでの処理が中心的になるため、そのハードウェア特性の差は広がる可能性が高い。しかしながら、両者において、単にハードウェアだけでなく、推論から学習、更にそのほかのデータ処理の活動で共通かつ広いソフトウェアのエコシステムが提供されていく必要があり、ハードウェア一辺倒の研究開発では実用的な普及はおぼつかないであろう。

更に長期的には、従来のノイマン型コンピュータと異なる計算原理である非ノイマン型コンピュータや、デジタル計算ではなくアナログ計算可能なデバイスを利用するニューロモーフィックコンピューティング、量子計算等の発展も念頭に置いておく必要がある。ニューロモーフィックコンピューティングのデバイス開発では、欧州のHuman Brain Projectでの「SpiNNaker」や「BrainScales」、米国のIBMの「TrueNorth」などが先行している。近年では各国で研究が進み、我が国のメーカーもいくつか開発に乗り出しているが、実用的な学習への適用はまだ基礎研究段階である。

国立情報学研究所（NII）やNTT、富士通などは量子計算の研究開発を実施しており、そのなかでも量子アニーリングは数理最適化の分野においていくつかの成果が報告されているが、実用的な機械学習への適用は未知数である。これらのデバイスでも、基礎研究段階では性能が重要なパラメータであるが、応用段階ではアルゴリズムやソフトウェアのエコシステムの手厚いサポートが重要になってくると考えられる。

1.7.2 ディープラーニングで要求される演算の基本

ディープラーニングの学習・推論に使われる演算のうち代表的なものとして、フィードフォワード（feed forward）計算（前向き計算）とバックプロパゲーション（back propagation）計算（後ろ向き計算）のそれぞれについて示す。

※4
Flops（Floating-point Operations Per Second）。1秒間に実行可能な浮動小数点演算の回数。

※5
“AI Bridging Cloud Infrastructure（ABCI）.” 産業技術総合研究所ウェブサイト <http://www.itri.aist.go.jp/events/sc2016/pdf/P06-ABCI.pdf>

1.7.2.1 フィードフォワード計算（前向き計算）

(1)全結合層

全結合層は、各ニューロンが前のレイヤーの全てのニューロンに、独立したシナプスで接続されているレイヤーである。全結合層のフィードフォワード計算は、本質的に行列（パラメータ・重み)、ベクトル（活性）積である。この計算を効率的に行うためには、複数のサンプルに関する活性ベクトルを行方向に連結して行列、行列積として計算する方法が広く用いられており、この計算は線形代数演算のAPIであるBasic Linear Algebra Subprograms（BLAS）のGEneral Matrix Multiply（GEMM）カーネルを用いて行われることが一般的である。

(2)畳み込み層

畳み込み層は、各ニューロンが前のレイヤーの近傍とパラメータを共通するシナプスで接続されているレイヤーであり、主に画像処理（2次元）で用いられる。畳み込み層はレイヤーの境界の扱いやフィルタのスライド幅（stride）に関して様々な変種が存在する。畳み込み層のフィードフォワード計算に関しては、計算内容を変えない範囲で様々な計算アルゴリズムが存在する。

- GEMMを用いる手法：
  パラメータと前のレイヤーの活性の畳み込みをベクトル同士の内積と解釈することで、GEMMを用いて畳み込みを行うことができる。行列積計算を行うためには活性を行列の形状にコピーする必要があるが、十分に大きい行列サイズの場合はGEMMの実行時間が支配的となる。一方でこの行列は重複する値を多数含むため、メモリを圧迫する可能性がある。この手法ではよく最適化されたBLAS実装を用いることで高効率な計算を行うことができる。
- ウィノグラードのMinimal filtering algorithm:
  「Minimal filtering algorithm」[1]は、入力サイズ・畳み込みサイズによって定まる定数行列を用いて適切に加減乗除を行うことにより、乗算回数を入力サイズとフィルタサイズの和に比例する計算量で1次元の畳み込みを行うアルゴリズムである。これを2次元に拡張することで、ナイーブな実装と比較して少ない乗算回数で畳み込み層の計算を行うことができる。このアルゴリズムは十分に大きいフィルタサイズが必要なFFT（Fast Fourier Transform）と比較して、フィルタサイズが小さい場合に特に有効である。
- FFTを用いる方法：
  二つの関数f, gについて、フーリエ変換をFとすればF（f∗g)=F(f)F(g）が成り立つ（ただしf∗gは2関数の畳み込み)。よって畳み込みニューラルネットワーク（CNN）の畳み込みについても入力とフィルタを離散フーリエ変換し、積を逆変換することで畳み込みを行うことができる。
- cuDNNのアルゴリズム：
  NVIDIAが開発するディープラーニング向けライブラリであるcuDNN[6]では、前述のアルゴリズムを含めた複数の畳み込みアルゴリズムが実装されており、ユーザが選択できるようになっている。これにより、ユーザが（多大なメモリを使用する）GEMMによるアルゴリズムよりも、高速かつ使用メモリ量が小さいアルゴリズムを使用できる余地があるとしている。

※6
GPUに特化したプログラミング言語であるCUDA (Compute Unified Device Architecture) を用いて書かれた深層ニューラルネットワーク (Deep Neural Network; DNN) 用のライブラリ。

1.7.2.2 バックプロパゲーション計算（後ろ向き計算）

ディープラーニングは、与えられたデータセットに対して深層ニューラルネットワーク（DNN）のパラメータを最適化する最適化問題に帰着される。パラメータの最適化には確率的勾配降下法（Stochastic Gradient Descent; SGD）が最も広く用いられる（数式による表現は後述のコラムを参照）。

SGDではパラメータの更新ごとにランダムに選択された少数個のデータを用いるため、1回の更新あたりの計算量を低く抑えることが可能である。大規模なデータを学習（特に教師あり学習）する場合には、SGDの利用が標準的である。一方、バッチ（データセット全体）を使った最急降下法では、1回の反復に必要な計算量が多大となり、大規模データの学習には現実的ではない。よって、SGDを用いることでほどよい反復あたりの計算量と収束性で学習を行うことができる。

バックプロパゲーションとは、DNNの出力に対する誤差を出力レイヤーから順にフィードフォワードとは逆方向に伝播させることで、各パラメータに対するコスト関数の勾配を計算する手法である。

＊DNNのパラメータθを最適化する最適化問題

$$\theta^* = \mathrm{argmin}_\theta \sum_{i \in D} E_\theta(z^i, t^i)$$

ただし$D=\{z^i, t^i\}$はデータセット（データz^iとラベルt^iの組の集合）、E_θはコスト関数であり、DNNの出力$DNN_\theta(z^i)$とラベルt^iの何らかの距離として定義される。

＊パラメータの最適化に用いられる確率的勾配降下法SGDの反復式

$$\theta^{(t+1)} = \theta^t - \eta \sum_{i \in M^{(t)}} \frac{\partial E_\theta(z^i, t^i)}{\partial \theta}$$

ただし$M^{(t)}$はtステップ目でデータセットからランダムに選択された部分集合（ミニバッチ）、$\eta>0$は学習係数である。

1.7.3 学習用のインフラストラクチャと計算デバイス

1.7.3.1 ディープラーニングによるAIシステムの背景

ディープラーニングによるAIシステムの発展を、画像認識タスクを例に説明する。ILSVRC（ImageNet Large Scale Visual Recognition Competition）[2]は、スタンフォード大学のコンピュータビジョン研究グループを始めとした専門家チームにより2010年に開設された、静止画からの一般物体認識の認識性能を競うコンペティションであり、この目的の為に研究用のデータセットが公開されている。訓練用データセットには1,000クラス、128万点のデータサンプルがあり、この規模は過去に公開された画像の教師ありデータセットの中では、最大規模のものである。

2012年のILSVRCにおいて、当時トロント大学（カナダ）のアレックス・クリジェフスキー氏（現在Googleに在籍）、イリヤ・スツケヴェル（Ilya Sutskever、現在OpenAIの研究ディレクターを務める）氏、ジェフリー・ヒントン氏の三氏の研究チームが、深層構造を持つニューラルネットワークを用いて認識精度におけるブレークスルーを達成したことは、有名である。1位のトロント大学チームと2位の東京大学チームの認識率の差は10%近くもあり、これがニューラルネットワークの「復活」を印象付ける結果となった。

ブレークスルー以前の典型的な解法は、特徴抽出部と識別部を使ったものであり、前者は経験的に設計され、後者は機械学習により最適化されるものであった。特徴抽出はコンピュータビジョンにおける主要な研究領域の一つであり、アプリケーションドメインにおける事前知識、静止画や動画、処理負荷などの観点から数多くの手法が提案されてきた経緯がある。一方、ニューラルネットワークでは特徴抽

出と識別は統一的に扱われており、特徴抽出部自体がパラメトリックに最適化される点に特色がある。更にDNNでは、特徴抽出部の多層化によってモデルの表現能力が高まり、多くの場合更に認識性能を引出すことが可能となる。[3]

ILSVRCの開設前は、PASCAL[7]と呼ばれるデータセットが一般物体認識の研究で広く使用されていた。このデータセットはILSVRCと比較してサンプル数が2桁少ないが、ここに興味深い研究報告がある。PASCALにおいては、特徴設計型の識別器のほうがDNNに比べて認識性能が良好であると報告されている[4]。つまり、ディープラーニングは小規模データセットにおいてはそれほど認識性能が出ず、大規模データセットが与えられて始めて秀逸な認識性能を獲得することができると考えられる。この意味でILSVRCの登場こそがDNNの“再発見”の最大の要因の一つと言える。

ディープラーニングの研究はブレークスルー以降爆発的な発展を見せ、2016年のILSVRCでは、クリジェフスキー氏らの達成した認識精度を更に10%以上上回るレベルに到達している。近年では特徴抽出過程の学習一般を扱う「表現学習」と呼ばれる分野の国際学会が創設されるなど、分野の発展に拍車がかかっている。

ネットワーク構造は複雑化の一途を辿っており、バイパス構造を持っているものや、ネットワークの途中でコスト関数が定義されているものなど、様々な構造が提案されてきた。現在ではDNNとは「誤差逆伝搬が可能な任意形状の計算グラフ」とのとらえ方が一般的になっており、いくつかのディープラーニングのソフトウェアプラットフォームもこの考えに立脚している。

画像処理だけでなく、人間を超える処理の可能性を強く印象付けたのはDeepMind（英国）が2015年に発表した「AlphaGo」である。それまでは、AIはチェスや将棋では人間のプロに勝利していたが、囲碁はあまりにも対局の場合の数が多くかつ複雑で、従来から用いられていた伝統的なAIの局面の探索手法（α-β探索法など）では、最新の計算機を用いても計算量が爆発するとみなされていたからである。

しかしながら、DeepMindは、2015年に探索を行う際の判定関数にディープラーニングネットワークとモンテカルロ木を用い、更にプログラムを相互に数千万回対戦させる強化学習を行うことによって、世界最強の棋士の一人であるイ・セドル（Lee Sedol）氏との対局で4-1で圧勝し、人間を上回る囲碁プレイヤーAlphaGoを作り上げた。その技術概要は『Nature』誌にも掲載され[5]、対局のネット配信とともにマスコミが大々的に宣伝し、AIの新時代として大いに注目を浴びることになった。

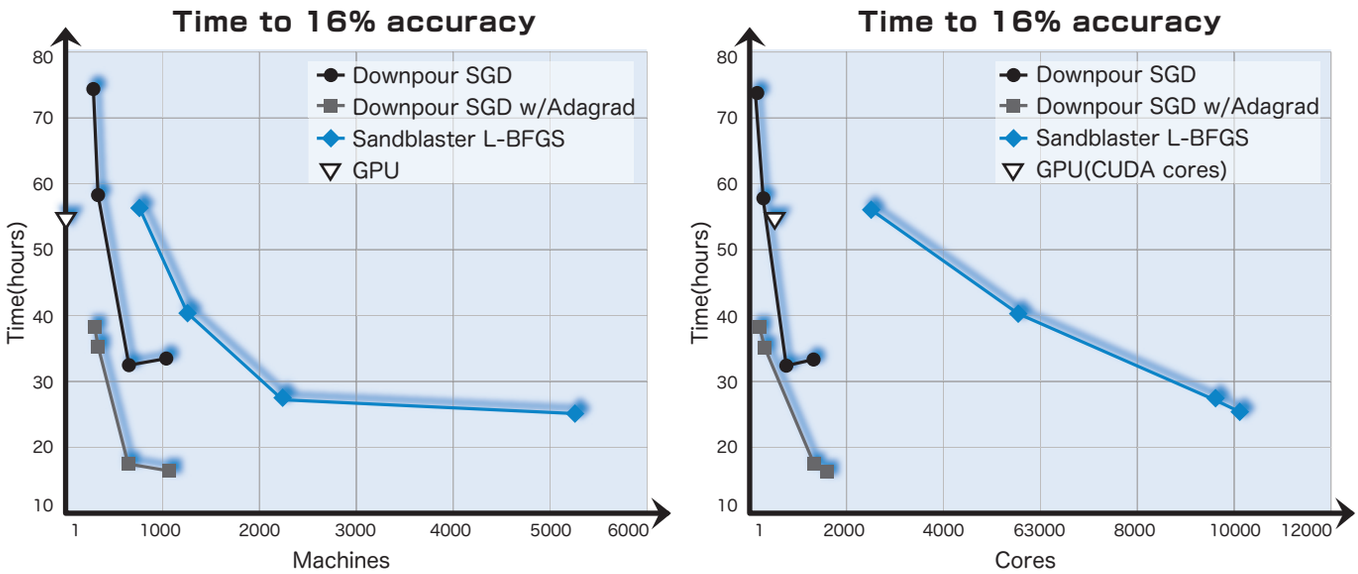
AlphaGoの勝因は二つある。一つは上記のようにディープラーニングの最新技術を適用したことであるが、二つ目は、それを可能にする十分な計算資源を確保し、有効活用したことである。画像処理同様に、Google Cloud Platform上に配備された数百のGPUを潤沢に用いることにより、複雑なディープラーニングネットワークによって駆動されるAIプレイヤーの莫大な対局数による強化学習に成功した。

ILSVRCではGPU中心のHPCによる加速が、それより早くから注目されていた。当初、Googleはディープラーニングの処理プラットフォームとして、自社のCPUファームを用いていたが、**図36**にみられるように、並列化による強スケーリング[8]時のスケーラビリティの達成が非常に困難であることが判明した。これは後に述べるように、データ並列の手法によって並列加速された学習を行う際に、各ノードで計算された新たな勾配の集約演算（collectives）が相対的かつ急速にオーバーヘッドになるからである。これを解決する一つの方法は、演算密度が高いだけでなく、内部のバンド幅も高いGPUを用いることであった。

このように、ISLVRC及びDeepMindやAlphaGoの成功により、GPUを中心とした高性能計算・スーパーコンピューティング技術によるAIの新たな展開が明らかになった。それとともに、クラウドでは

※7
“The PASCAL Visual Object Classes Homepage” University of Oxford Website <http://host.robots.ox.ac.uk/pascal/VOC/>

※8
問題のサイズを変えないでプロセッサ数を変えた時の計算時間の評価。



■図36　計算機ノード数・コア数を増減させたときの認識精度16%達成にかかる時間[9]

HPC技術のIDC（Internet Data Center）への導入が急速に始まり、逆にHPCでは計算シミュレーション中心の世界から、ビッグデータやその解析に対する機械学習の適用における数値計算アルゴリズムやシステム技術の適用に関心が高まり、結果として更にディープラーニングを中心としたAIの加速や大規模並列化へ向けたアルゴリズムプロセッサシステム・計算インフラストラクチャ等の研究開発が世界中で盛んになってきている。

このような世界的なトレンドに対し、我が国は遅れを取っていることは否めない。機械学習に関するアルゴリズムの基礎研究や、応用事例は様々あるものの、Google、Amazon、Facebook、Microsoft、Baidu（百度、中国）などの、米中のいわゆる「AI Giants」と比較し、それに匹敵するインフラストラクチャや、システムハードウェアやソフトウェアに対する研究は立ち上がりが遅れ、かつ、HPC分野からの関心も未だ薄い。しかしながら、近年は一部それらを解決する動きが、東京工業大学のTSUBAMEや産総研のABCI、富士通DLU（Deep Learning Unit）など、産官学でも進められており、これらの努力を更に拡大していくことがAI後進国にならないためにも大変重要である。

1.7.3.2 学習向けの高性能インフラストラクチャ・スーパーコンピュータ及びGPUの台頭

元来グラフィックス計算用のプロセッサとして開発されたGPUをHPCに転用するGPGPU（General Purpose computing on GPU）技術はこの10年で一般化した。

グラフィックス処理においては、ピクセル値などの大量のデータ列・パラメータに対して同一の計算を適用することが多く、GPUの設計もそれを実現するために、汎用計算用のCPUと比較してSIMD（Single Instruction Multiple Data）的な処理に特化したものとなっている。そのため、CPUと比して複雑な分岐処理などの性能には劣るが、大量のデータに対して同一の演算を行う並列性の高い処理については、同時代のCPUの5～10倍の性能を示す。

GPUで動作するプログラムを作成するには、従来はCUDA（Compute Unified Device Architecture）やOpenCL（Open Computing Language）などの、GPUに特化したプログラミング言語を用いる必要があったが、近年ではOpenACC（Open Accelerators）などの、C言語プログラムにディレクティブと呼ばれる記述を付加するだけでGPU用の並列化を達成させる言語拡張や、GPUで行わせる

※9
文献[6]より引用。

■表4　国内外の主なGPUスーパーコンピューター覧

導入年	コンピュータ名	GPU 及び搭載数	理論性能値 （HPC 向け倍精度）	理論性能値 （機械学習向け短縮精度）
2010	TSUBAME2.0	M2050 × 4224	2.4 PFlops	4.8 PFlops
2013	TSUBAME2.5	K20X × 4224	5.7 PFlops	17.1 PFlops
2013	Titan	K20X × 18688	27 PFlops	80 PFlops
2016	Piz Daint	P100 PCIe × 4500	15.99 PFlops	82 PFlops
2016	DGX Saturn V	P100 SXM2 × 1000	4.90 PFlops	19 PFlops
2017	TSUBAME3.0	P100 SXM2 × 2160	12.15 PFlops	47.2 PFlops

処理をまとめたライブラリインターフェースを用いて間接的にGPUを利用するなどの、言語習得コストの低いプログラミング方式が実用できるレベルにまで発展している。

GPU技術の一般化はスーパーコンピュータを用いる科学技術計算にも当てはまり、**表4**に示すとおり日本及び海外のスーパーコンピュータにもGPUを多数搭載したものが現れ、流体力学、分子動力学を始めとした様々なシミュレーションプログラムがGPUを用いて実行されている。国内では、東京工業大学が2010年に運用開始したTSUBAME2.0が導入時からGPUを主たる計算資源として設計されたスーパーコンピュータの先鞭をつけ、2013年にTSUBAME2.5としてGPUのアップグレードを行い、2017年8月からは後継機TSUBAME3.0の運用が開始される予定である。

従来、HPCアプリケーションでは、倍精度浮動小数点数（64bit、double）を用いた演算が行われてきたが、近年では反復法を用いるものなど多くのアプリケーションにおいて、単精度浮動小数点数（32bit、single）でも実用上十分な精度の解が得られることが分かってきた。GPUは機種によるものの、倍精度演算の2倍以上の速度で単精度演算を行うことができ、演算とともにプログラム実行速度のボトルネックとなるメモリ転送速度の面でも、通信量が半分になる単精度浮動小数点数の採用は有利となる。

この傾向は近年更に先鋭化し、2016年に発売された最新世代のGPUであるNVIDIAの「Pascal P100」では半精度浮動小数点数（16bit、half）の演算がサポートされ、倍精度の4倍、単精度の2倍の速度で演算することができるようになったため、ディープラーニングに代表される浮動小数点演算の精度を要求しない機械学習アプリケーションの実行が更に高速化できるようになった。これらの性能向上とともに、Google、Amazon、Facebook、Microsoft、Baiduなどの、いわゆるクラウドの「AI Giants」でもGPUの採用・配備が進んでいる。

このように、主にシミュレーションサイエンスに活用されてきたGPUスーパーコンピュータであるが、AI分野においても、前節に示されるとおり、行列積やFFTなど、GPU処理に向いたアルゴリズムが用いられており、近年複数ノードを用いて大規模なネットワークの学習を並列化して実行する例が多数報告されている。また、ネットワークパラメータの最適化のために、複数のパラメータにおける学習を多数のノードで並行に計算するアンサンブル並列学習や、それらを組み合わせた実行も行われている。これらに関しては後述する。

1.7.3.3 ディープラーニング専用プロセッサ

現状では、バックプロパゲーションSGDによるディープラーニングは非常に長い時間と莫大なデータセットの供給が必要であり、かつDNN自身もギガバイト単位の多くのメモリを要求する百以上にも上

※10
エッジデバイスとは、クラウドに、あるいは大規模なサーバのようなネットワークの中心に接続する、末端のデバイスのこと。IoTにおける個々のモノやセンサ類にあたり、処理能力や消費電力は小さい。

る深い層で構成されていて、計算資源や電力の乏しいエッジデバイス[10]における学習は現実的ではない。どちらかと言えば、学習フェーズは先にあるようなHPCから派生してきた大型のチップが装備されている、高性能クラウドやスーパーコンピュータで行うのが主流となっている。

既に述べたように、密行列演算やGEMMの非常に高い加速性能から、ディープラーニングにブレークスルーを最初にもたらしたハードウェアはGPUであった。現状では、ほとんどのディープラーニングのワークロードはGPUによってなされていると言っても過言ではない。しかしながら、いくつかのメーカーがディープラーニングに向いた、あるいはディープラーニングに特化したクラウドスーパーコンピュータ向けの大規模チップを研究開発しており、2017年後半近辺から市場に出回り始める観測である。これらにより、GPUの寡占状態が改まり、競争によって更なる技術革新が起こることが期待されている。

しかし、クラウドやスーパーコンピュータのサーバ向けのチップは、単に性能を上げるだけでは不十分であり、大規模データセンター運用に供するためのRAS（Reliability Availability Serviceability)、安定性やリモート制御、更には長めの製品サイクルや継続性など、様々なエンタープライズ向けの機能が必要であり、そもそもの複雑さと相まって、通常のITと同様、組み込み系と比べて遥かにハードルが高い。

更に、CaffeやTensorFlow等の数多くのディープラーニングのフレームワークや関連ツールが、そのチップで汎用かつ最高性能で動くように、ソフトウェアの開発やチューニングから維持サポートまで、多くの関連開発が必要である。つまり、エンタープライズ系のCPUと同様に、単に「動く」ものでは駄目であり、その開発は、時間・経験・人手が豊富な大手ITベンダーに限られている。

以下、現在アナウンスされているプロセッサをいくつか紹介する。

- ディープラーニング向けGPU:

  NVIDIAは2016年に、機械学習に向けたハードウェア特性を備えた大型のGPUである「Pascal P100」を発表・出荷し、現在多くのクラウドやスーパーコンピュータで大規模な採用が始まっている。Pascalの機械学習向けの最大の特徴点は、HPCでは通常用いないFP16という16ビット幅の浮動小数点形式のハードウェアのサポートであり、これによりGEMMの理論ピーク性能が従来用いられていた32ビットの約2倍であるチップあたり21.2 TFlopsと、大変高いものになっている。また、HPCと共用の新たな特徴として、後述の並列化時のチップ間の高速通信を実現するNVLinkが装備され、マルチGPUの学習が従来と比較して大幅に加速される。NVIDIAは更に、Pascal P100を8機装備した機械学習専用サーバであるDGX-1を開発し、多くの研究所や開発の現場で小～中規模の学習プラットフォームとして用いられている。

  NVIDIAは、2017年末～2018年初頭に発表する次期GPUの「Volta」では、より機械学習向けの専用機能を発展させ、全体的な高速化とともに、様々な縮退精度（16 bit、8 bit）のサポート、テンソル演算の高速化命令などを実現すると表明している。また、大規模GPUではライバルとも言えるAMD（米国）も、Vegaシリーズの機械学習向けの「Radeon Instinct」GPUを発表し、そのFP16性能では25 TFlopsと、P100の性能を上回ると主張している。

- Intel Xeon Phi - Knights Mill及びLake Crest:

  IntelはHPC向けに高性能メニーコア型CPUである「Xeon Phi-Knights」シリーズを数年来開発しており、最新の「Knights Landing」（KNL）プロセッサは米国エネルギー省サンディア国立研究所の「Trinity」や同ローレンス・バークレー国立研究所の「Cori」、我が国では東京大学「Oakforest-PACS」などの最新のトップスーパーコンピュータに採用されている。それをベースに、Intelは機械学習向けに改良した「Knights Mill」（KNM）プロセッサを2016年に発表し、2017年

中に出荷するとしている。KNLと比較して、KNMはHPCでは多用されるが機械学習ではほとんど用いられない倍精度浮動小数点演算性能を縮退し、その代わりに用いられる単精度演算を大幅に強化している。更に、FlexPointという（FP16ではない）新たなディープラーニング用の演算形式を採用し、性能面ではP100を上回る、としている。

また、Intelは2016年にNervana Systems（米国）を4億ドルで買収し、開発されていたディープラーニング専用チップを製品ラインアップに加えるとしている。GPU同様比較的汎用性が高いKNMと比較して、その第一弾である「Lake Crest」はより機械学習向けの縮退精度のテンソル演算に特化していると言われており、その専用化によりKNM比でもより高い学習性能が実現されるとIntelは主張している。

- FPGA:

組み込み系と同様に、FPGAを大規模ディープラーニングに用いる試みも盛んになってきている。特に、Intelに買収されたAltera（米国）や、そのライバルのXilinx（米国）は、サーバプロセッサ並みの巨大なデータセンターやHPC向けのFPGAを開発しており、更に、最新版の「Altera Stratix 10」や「Xilinx Virtex UltraScale+」では、内蔵のハードマクロ演算ユニットであるDSP（Digital Signal Processor）を大量に装備し、ディープラーニングにも効率よく対応できるとしている。

FPGAはDNNで用いられるビット長を任意に調整したり、あるいはハードDSPを使う場合でもデータの流れを最適化したりするなどの手法で、電力性能比でGPUを上回る可能性が学会などで報告されている。しかしながら、FPGAでのプログラミングのツールチェインは、やはり通常のCPUやGPU程は整備されておらず、かつ、推論と比べて複雑なディープラーニング系のアルゴリズムを有効かつ気軽に載せるのはまだまだチャレンジが大きい。

そのようなソフトウェア上の高いハードルはあるものの、MicrosoftのAzureなどの一部のクラウドIDCでは、後述するようにFPGAを用いた学習を行っており、今後が注目される。

- 富士通DLU:

我が国では富士通が、ディープラーニング専用プロセッサである「DLU」を開発している。詳細はまだあまり明らかにされていないが、京コンピュータで培ったHPC関連の様々な技術を投入し、現状のGPUと比較して10倍の電力効率を得るとしている。その登場は2018年末とされているが、富士通が開発中のAI向けの統合フレームワークである「Zinrai」に統合される予定であり、今後注目である。

- Google TPU:

Googleは2016年に自社開発のディープラーニングASICである「TPU」を発表した。既にGoogleのIDCに相当数が配備され、実際のアプリケーションに供されていると発表している。その内容は、8ビットの行列積を行うハードウェアがシストリックアレイ[11]状に約16,000個並んでいるという、演算エンジンとしては比較的クラシックな構成になっている。その性能は高いものの、精度の低さなどの要因により、TPUは推論専用であり、学習エンジンとして用いるのは困難である。それでも自社IDCで商用向けに大規模運用を行っている点は大変評価でき、今後学習向けのプロセッサも開発してくる可能性はあるであろう。ただし、この論文での性能評価に関し、上記の最新のPascal P100

※11
小さい処理を行うプロセッサを多数接続して、近接プロセッサ間において大量のデータ交換ができるようにした構造。

GPUを用いるとそのほうが推論性能は約倍で、かつ学習も可能であるとNVIDIAが声明を発表しており[12]、これらからも専用チップ対GPUの競争の激しさが鑑みられる。

また、上記のほか、NECは大阪大学、情報通信研究機構（NICT）脳情報通信融合研究センター（CiNET）、理化学研究所生命システム研究センター（QBiC）と連携して、脳型コンピューティングシステムの開発を行う研究所を開設したほか[13]、東京大学とも共同研究を開始している[14]。

1.7.3.4　ディープラーニングの大規模並列化

（1）大規模システムの背景と並列化

ディープラーニングの加速化は個々のプロセッサの進化だけでは不十分である。実際、新時代の百層以上にもわたる“深い”ネットワークや、データセットの増加や多様化による次元の増加での計算需要の増加は、プロセッサ側のムーアの法則[15]やアーキテクチャ革新による進化を遥かに上回っている。また、ディープラーニングの適用領域の広範化や、高精度化の要求による学習及び推論用のデータの増加や解像度の進化も、ビッグデータ処理能力の大幅な需要増加を意味し、現状でも何ペタバイトにも上るデータを柔軟に扱うには、小規模システムでは困難である。

このような劇的な需要増加に対して、既にGPUの採用が進んでいることは述べたが、総合的に鑑みればハードウェアアーキテクチャとして最適であるものは、現代の超並列アーキテクチャ型のスーパーコンピュータであることは自明であろう。The Top 500[16]やGraph500[17]での世界の上位ランクのトップスーパーコンピュータは、数千～数万個のマルチコアやメニーコア型のプロセッサを数十～数百ビット／秒の超高速ネットワークで密結合し、更にテラバイト／秒の超高性能のI/Oを備えている。AI用クラウドで採用が進むGPU自身も、グラフィックス以外の計算用途への適用はエンジニアリングワークステーションやHPCが最初であり、現在での多くのトップスーパーコンピュータがGPUやその派生型のアーキテクチャを採用している。クラウドもAIへのシフトが起こるにしたがい、基本的にはスーパーコンピュータをそのインフラストラクチャ内に構築しており、我が国のさくらインターネットの高火力コンピューティング[18]もその一種である。

HPCにおける速度向上のテクノロジーの最も基本となるのは処理の並列化である。よって、プロセッサのいかんに関わらず、その内部の並列化と、プロセッサ間の並列化を効率よく果たす必要がある。現在のHPCでは、トップスーパーコンピュータにおける全体の並列性は数百万以上に上り、それを効率よく生かせるアルゴリズムやシステムの研究が過去より中心的な課題として盛んに行われてきた。機械学習、特にディープラーニングの学習フェーズでの並列化も例外ではなく、近年のその隆盛により、多くの超並列化による高速化の研究開発が行われている。

※12
“AI Drives the Rise of Accelerated Computing in Data Centers.” NVIDIA Website
<https://blogs.nvidia.com/blog/2017/04/10/ai-drives-rise-accelerated-computing-datacenter/>
※13
NEC Brain Inspired Computing 協働研究所ウェブサイト
<http://nbic.ist.osaka-u.ac.jp/index.html>
※14
「NECと東京大学、日本の競争力強化に向け戦略的パートナーシップに基づく総合的な産学協創を開始　～第一弾、AIの共同研究・倫理/制度の検討・人材育成を推進～」NEC Website
<http://jpn.nec.com/press/201609/20160902_01.html>
※15
大規模集積回路（LSI）の将来予測（集積回路上のトランジスタ数が1.5年ごとに倍になる）
※16
“TOP500 Supercomputer Sites.” Top500 Website
<http://www.top500.org>
※17
“The Graph 500.” Graph500 Website
<http://www.graph500.org>
※18
「高火力コンピューティング」さくらインターネットウェブサイト
<https://www.sakura.ad.jp/koukaryoku/>

ただし、いかにスーパーコンピュータといえども、無制限の演算能力とメモリ空間、アクセス速度が得られるわけではなく、むしろ学習の大規模化に伴い、学習の演算自体ではなく、途中経過である学習パラメータの勾配情報等の通信処理が全体の実行時間の過半を占めるようになる。そのため、高性能計算におけるアプリケーション最適化と同様の方法論でアプリケーションのボトルネックを解析し、当該箇所のアルゴリズムの変更を含む最適化を行うことによって、学習全体の速度及びスケーラビリティの向上を行うことが必要不可欠である。

これらの中には、SqueezeNetのようにネットワーク中のフィルタやそこに通すチャネル数（色数）を削減することで演算数を削減する試みや、半精度浮動小数点数やブール値を含む低精度演算を用いることで、演算・メモリ使用量・通信量を大幅に削減する手法も含まれる。

また、スーパーコンピュータを用いて作られた大規模モデルは推論時にも多くの演算を要求するが、同等の推論をより小規模のネットワークで再現させるModel Distillationのような技術も今後重要になると考えられる。

（2）ディープラーニングの大規模並列化とその課題

ディープラーニングを行う上での課題の一つは、DNNの学習に極めて長い時間を要する点である。DNNの学習は主に計算律速[19]であり、単一GPUを用いた学習では数週間～数か月程度の時間を要する場合がある。また、最適なDNNの構造ハイパーパラメータは一般に自明ではなく、推論性能の良いDNNを得るためには長時間の学習と手動によるDNNの構造・ハイパーパラメータを交互に繰り返すことが必要とされる。この問題を解決するために、スーパーコンピュータ技術を利用した様々な形態の並列化が試みられている。

第一の並列化は、プロセッサ内部の並列化である。既に述べたように、ディープラーニングの計算カーネルは密行列演算、Winograd法、FFT法に大別されるが、それらはHPCの分野では広範囲なアプリケーションにおいて並列化・高性能化の研究や実装が長年行われてきた分野である。ただし、いくつかディープラーニングに特化するべき部分があり、例えばCNNの行列演算では多くのHPCアプリケーションの場合と異なり、幅が狭く長さが長いいわゆるTall and Skinny Matrix同士の密行列演算が畳み込み演算時に頻出し、その効率の良い並列演算アルゴリズムが開発競争の最先端となっているのが現状である。

次に、プロセッサあるいは計算ノードをまたぐディープラーニングの分散並列化手法は、主に「データ並列」と「モデル並列」に分類される。ここでの「データ」とはデータセットに含まれるサンプル（例：画像認識タスクにおける画像）やそれらを処理することによって得られるデータ（例：フィードフォワード計算で生じる活性等）のことを表し、「モデル」とはDNNそのものを表す。

- データ並列：
  複数の計算機が同一のDNNのパラメータを持ち、異なるデータについて並列に計算を行う手法。各データサンプルの処理によって生じるデータは計算モデル上では独立であるため、これらを通信する必要はないが、DNNのパラメータを同期するための通信が必要となる。
- モデル並列：
  複数の計算機が同一のデータを分割して持ち、異なるモデルの箇所について並列に計算を行う手法。同一のデータを複数の計算機が使用するため、高頻度な通信が必要となる。一方で、計算機同士でDNNのパラメータを重複して持つ必要がないため、パラメータ数が大きく単一の計算機のメ

※19
律速はボトルネックを意味し、つまり計算律速とは計算能力がボトルネックになっていることを表す。

モリに入らない場合にも有効である。

データ並列とモデル並列のどちらがより有効であるかどうかは、DNNの構造により異なる。一般に、畳み込みレイヤーはパラメータ数が小さいためにデータ並列が適しているのに対し、全結合レイヤーはパラメータ数が大きいためにモデル並列が適しているといわれる。また、GPUクラスタ（スーパーコンピュータ）で学習を行う場合は、GPU内では多数の内蔵コアを用いてモデル並列を利用し、GPU間では後述する集団通信等を用いたデータ並列を用いるなど、これら二つの手法を併用することが一般的である。

これらの並列化手法を応用した学習手法として以下の例が上げられる。

- パラメータサーバ：

  パラメータサーバとは、学習途中のDNNのパラメータを専用のサーバ（群）で一元管理する手法、又はそのサーバのことを指す。学習に必要な演算を行う計算機（ワーカー）は更新量を計算してパラメータサーバに送信し、対してパラメータサーバは受信した更新量を用いて内部に持つパラメータを更新してワーカーに返すことにより学習が進行する。後述する集団通信による手法と異なり通信がワーカー、パラメータサーバ間の一対一通信のみで成立することが利点として挙げられる。また、パラメータサーバはパラメータを複数に分割して分散key-valueストアの要領で複数のサーバに分散させることが可能である。パラメータサーバを採用した例としてはGoogleの「Dist Belief」、Microsoftの「Project Adam」が挙げられる。
- 集団通信による手法：

  ミニバッチを計算機に分散させてデータ並列で学習を行い、別々に計算された更新量の総和を集団通信によって計算する手法。ミニバッチSGDではコスト関数が各サンプルの総和又は平均として定義されているため、計算機ごとに計算された更新量の総和又は平均を計算することで、ミニバッチSGDの計算モデルに則った更新が可能となる。ここでの通信では主にall-reduce（全プロセスのデータをリダクションし、結果を全プロセスに複製する通信）が用いられる。この手法を採用した例としてはBaiduの「Deep Image」や「Deep Speech 2」が挙げられる。

  集団通信の計算精度は通信速度向上のために低く設定されることがある。例えば、Microsoftが開発しているディープラーニングフレームワークであるCNTKでは、一つの通信値（勾配）を1ビットで通信する手法が実装されている。この手法では、通信値を符号によって0又は1で量子化し、復号する際にはパラメータ行列の列ごとの（正値又は負値の）平均値を用いる。また、量子化によって生じる量子化誤差を次回の通信に持ち越すことで、量子化誤差が累積しないようにする。このような精度を下げた場合の通信が学習に与える影響は、計算の精度を下げた場合と同様に自明ではない。

パラメータサーバや一部の集団通信による手法では、非同期な学習が用いられている。非同期学習とは、パラメータの更新やそれに伴う通信中にも、古いパラメータを用いてフィードフォワード計算やバックプロパゲーション計算を行うことを指す。非同期学習は厳密にはSGDの更新則にしたがっておらず、同期学習と比較して収束性や学習終了後の推論性能、再現性が悪化することが予想されるが、一方で計算と通信のオーバーラップにより学習が高速化することが期待される。

（3）超並列化に向けた課題

ディープラーニングにおいて超並列化を阻害する一番の原因は、SGDが逐次的な計算と更新を必要とする点である。これにより、同期学習での並列数は（ミニバッチサイズ）×（モデルの分割数）以下に

制限される。更にデータ並列、モデル並列、非同期学習（パラメータサーバを含む）のいずれの並列数を増加させる場合でも以下のような問題が発生する。

- データ並列：
  多大なミニバッチサイズを用いる場合、収束性及び収束後の推論性能が悪化することが指摘されている。データ並列の利用によりミニバッチサイズが増加するとエポックあたりの学習速度は増加するものの、エポックあたりの収束速度が低下し、結果として学習速度が低下する可能性がある。
- モデル並列：
  一般に一つのレイヤーのフィードフォワード、バックプロパゲーション計算では前（後）のレイヤーの全フィルタにまたがるデータを必要とすることから、モデル並列を用いる場合はレイヤーごとに全対全通信が必要となる。近年では100層を超えるDNNの学習手法が確立されていることから、このようなDNNを学習する場合は通信時間の増加が課題となる。
- 非同期学習：
  非同期学習ではコスト関数の勾配の計算途中にパラメータが更新されてしまい、勾配が古くなる「staleness」と呼ばれる現象が発生する。非同期学習ではワーカー数が増加すると通信遅延が増加してstalenessが増加することから、収束性能が悪化する。

一方、ディープラーニングの高速化を考える上で重要なことは、ディープラーニングでは厳密・高精度な計算は必ずしも必要とされないという点である。例として非同期学習は厳密には同期学習とは計算内容が異なるものの、特定の条件では同期学習と遜色ない学習がより高速に行えることが報告されている。また、計算精度を単精度よりも低い精度で行う試みも盛んに行われている。

これらの手法は計算性能を向上させる半面、学習性能への影響については以前未知である。よって、ディープラーニングを今後超並列化するためには、古典的な科学技術計算で必要とされてきた計算性能（1エポック[20]の学習を行うために必要な計算時間; time-to-epoch）と同時に学習性能（満足な条件に収束するために必要なエポック; epoch-to-convergence）の両者を同時に考慮する必要がある。これらのような性能と得られる解の精度を総合的に鑑みる枠組みをHPCでは一般的にUQ（Uncertainty Quantification = 不確実性の定量化）と呼ぶが、その適用による適切な性能モデル化も大きな課題である。

1.7.3.5　AI向け高性能インフラストラクチャ

HPCと異なり、AI向けの大規模インフラストラクチャの整備はむしろ民間のメジャーなパブリッククラウドベンダーのほうで始まった。スーパーコンピュータにおいてもシミュレーションの結果や観測データのデータ解析は重要なワークロードであり、一般的な統計的なデータ解析手法は多く適用されてきたが、特にAIに対する取組が民間に対して遅れた理由は定かではない。現状では、特に米国や中国において、パブリッククラウドベンダーがGPUや高速ネットワークなどのスーパーコンピュータ技術を取り入れたAI向けのクラウドインフラストラクチャを急速に拡充させていっている。

一方、スーパーコンピュータ側も従来のシミュレーション一辺倒から、データ解析やAIに少しずつ主軸を移し始めている。しかしながら、GPUや汎用のメニーコアのマシンはクラウド同様対応が進んでいるが、専用プロセッサで構成されているマシンでは、汎用のディープラーニングフレームワークが必

※20
データセット全体を使った一回分の学習のこと。

ずしも動作するわけではなく、ハードルが現状ではまだまだ高い。これらが公的なディープラーニング中心のAI研究における阻害要因となっている。

HPCインフラストラクチャのように、世界のトップランクのスーパーコンピュータが1,000万プロセッサコア数を超え、公的な研究に処しているのに比較し、現状ではAI向けのインフラストラクチャは、計算及びデータに関して、民間のインフラストラクチャの方が豊富である。しかしながら、後者のアクセスはコストが非常に高く、すそ野の広い公的研究の阻害要因となっているのが現状である。

2017年3月現在、Google、AmazonなどのGPUの計算ノードのTFlops辺りの単価は月額2万円以上と高額であり、そのほかの付加コストも相まって、我が国の情報基盤センター等の同様の単価と比較すると数倍のコストがかかる。欧米のスーパーコンピュータ全般や、日本でもHPCI（High Performance Computing Infrastructure）などの全国組織から割り当てられるスーパーコンピュータの利用権は、審査ベースで基本は無料である。一方で、基礎研究ですら、商用基盤で競争力のある研究を行うと数千万円単位の金額がかかる状況とは、非常に疎遠であるといえる。いわんや、それらの基盤を所有するゆえに、競争力のある研究が行えているGoogleやBaiduなどの付属研究所は、AIや機械学習などの国際会議やジャーナルに多くの研究成果を発表しているのが現状である。

このような状況を打開するために、公的なインフラストラクチャの整備も行われ出している。特に、我が国では**表5**に示すとおり、国立のAIの3センターや、その協力機関などにおいて、AI向けの公的インフラストラクチャの整備が急ピッチで進んでいる。

■表5　我が国のAI専用・AI向け公的インフラストラクチャ

導入年月	システム名	システム概要	理論性能値 （ディープラーニング向け精度）	研究機関
2017年4月	AAIC （AIST AI Cloud）	NVIDIA Pascal P100 × 8 GPUサーバ × 50台	8.4 PFlops	産業技術総合研究所 AI研究センター
2017年4月	ディープラーニング 解析システム	NVIDIA DGX-1 （Pascal P100 × 8 GPU サーバ）× 25台	4 PFlops以上	理化学研究所 革新知能統合研究
2017年8月	TSUBAME3.0 ＋ 2.5 ＋KFC（HPC共用）	NVIDIA Pascal P100 × 2160 ＋ K20X × 4080 ＋ K80 × 168	65.8 PFlops	東京工業大学 学術国際情報センター
2018年3月	ABCI （AI 橋渡しクラウド AI Bridging Cloud Infrastructure）	未定（設計調達中）	130〜200PFlops	産業技術総合研究所 AI研究センター
参考 2017年	高火力コンピューティング	NVIDIA TitanX ＋ Pascal P100	不明	さくらインターネット

残念ながら、現状では**表5**に掲載した以外では、日本のスーパーコンピュータセンター並びにパブリッククラウドベンダーにおいて、AI向けのインフラストラクチャの発表を表明しているところは存在しない。公的なスーパーコンピュータセンターでは、新規の設計調達には時間がかかり、かつ既存のユーザベースを蔑ろにすることも困難で、計算とデータサイエンスが両立するようなストラテジーを長期にわたって積み上げてきた東京工業大学学術国際情報センター（GSIC）のような状況でない限り、即時の対応は困難である。

また、パブリッククラウドも同様であり、機械学習に供する大規模なインフラストラクチャは実質的にスーパーコンピュータをIDCに導入する必要があるが、電源供給・冷却・ネットワークの高密度実装、更には運用面において、対応が大変困難である。例えば、AI指向のスーパーコンピュータである東京工業大学TSUBAME3.0は、一ラックあたりの熱密度が最大61KWであるが、これは一般IDCのラック当たりの3〜6KWと比較すると10〜20倍にも相当し、このようなインフラストラクチャを通常のIDCに入れるのは困難である。

2018年3月稼働を目指して開発・整備が進められている、AIRCのABCI（AI橋渡しクラウド、AI Bridging Cloud Infrastructure）の一つのミッションは、この事態を打開することにある。ABCIの目標としては、①我が国に、ディープラーニングを中心とした計算とデータを両立させた、米国のパブリッククラウドベンダーに匹敵する130ペタフロップス以上の大規模AIインフラストラクチャを整備し、官民のAI研究者に提供すること、②そのようなインフラストラクチャの構築・運用のシステム技術を研究開発すること、③そのインフラストラクチャ技術を速やかかつ継続的に民間移転し、我が国のAI技術の全般的な研究開発のハブとして機能すること、などである。

そのなかでも、AIに向けたデータセンターの構築・運用の研究開発は中心課題であり、スーパーコンピュータで培われた高密度・高信頼・高性能・高効率の計算とデータ処理技術を、いかに安価かつIDCフレンドリーなインフラストラクチャとして技術移転していくかが、大きな目標となる。そのために、ABCIを開発しているAISTは、経済産業省産業技術環境局とともに、東京大学と「（仮称）グローバルAI研究拠点」に関する協定を締結し、東京大学柏キャンパスに新たなAI向けのデータセンターを構築し、日本の民間IDCのAI施設導入の模範的なショールームとする事業を進めている。いわば、TSUBAME3のようなスーパーコンピュータ専用の高密度実装技術を、IDC向けにほとんど性能を犠牲にしない形でコモディティ化し、全国のIDCへの速やかな普及を目指す予定である。

米国でも各研究機関のパブリッククラウドベンダーとの協業が進んでいるが、国立科学財団（NSF）、エネルギー省（DoE）などが2018年以降に向けて、AI・ビッグデータ指向のインフラストラクチャの整備を検討中である。

1.7.3.6　今後：HPCとの融合か離散か

以上、現状のAI向けのハードウェアのインフラストラクチャや技術進展、更には今後のロードマップを俯瞰した。ディープラーニングを中心としたAIの急速な進展は、単に分野内の研究開発だけでなく、HPCのシステム技術が多く転用され、それがAI向けにカスタマイズされ、アルゴリズムとのコ・デザインでもたらされていると言える。HPC向けの「エクサ」（10の18乗）フロップスのスーパーコンピュータが登場するのは2020～21年以降と言われているが、AI向けのスーパーコンピュータは、縮退精度を活用してより早く登場する見込みである。

例えば、米国オークリッジ国立研究所（Oak Ridge National Laboratory）の「Summit」スーパーコンピュータは2018年に稼働するが、最新のNVIDIA Volta GPUを心臓部分としたAI向けの性能は、800ペタフロップス以上と言われており、エクサ級と呼んで差し支えないであろう。更に、それらのマシン上の莫大なAIの能力を生かして、シミュレーションの一部をAIによるデータ予測で補完しよう、といった研究もいくつか始まっていて、AIとHPCの距離はアルゴリズムやアプリケーションレベルでも近づいているとも言える。

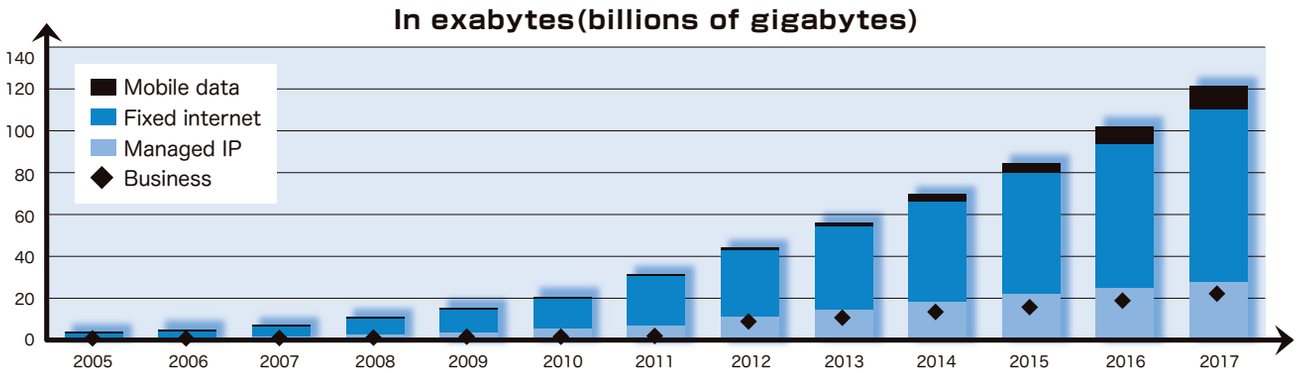
一方、今後このようなAIとHPCの協業が続くのか、あるいは分離するのか、という議論も存在する。例えばディープラーニングの精度要求を更に下げて高速化しよう、という研究は多くあるが、これら縮退された数値精度が他のHPCアプリケーションで活用できるかはまだ定かではない。同様に、更にデータフローがAIに特化された場合や、ニューロモーフィック計算のように、通常のHPCワークロードの計算モデルとしての適用性が定かではない技術が台頭してきたときに、それらのHPCとの融合は困難になる可能性もある。今後の高性能アーキテクチャの設計のためには、一般HPCと、よりマーケットが大きくなるAIが、どこが共通で、どこが異なるか、という技術的な探求が早急に求められていくであろう。

1.7.4 推論用のインフラストラクチャと計算デバイス

1.7.4.1 推論用のインフラストラクチャと計算デバイスのトレンド

推論時のインフラストラクチャや計算デバイスに求められる性能は、学習時とは大きく異なる。特に、ネットワーク上のデータのトラフィックから課せられる制約や、分野によっては処理の時間的遅れ（レイテンシ）に対する制約、使用可能なメモリ量に対する制約、消費電力に対する制約等が、学習時とは異なる形で要請される。

データトラフィックの観点からは、推論のために世の中に流通する全データをクラウド等で集中管理することは、非現実的であり、エッジ、あるいはフォグでのコンピューティングが重要となると考えられる。今後IoTの社会実装が進んだ場合、世の中に流通するデータ量が増大し、2020年には40ゼタバイトに到達すると考えられている。我が国のデータ流通量も、2005年の約1.6エクサバイトから2014年の約14.5エクサバイトまで、約9.3倍に伸びており、今後も増大トレンドは変わらないと考えられる。



■図37　インターネットトラフィックの拡大[21]

ロボット、自動運転等、即時的な応答が求められる応用分野では、情報処理に対するレイテンシに対する要求が厳しくなる場合が多い。例えば、Googleのデータセンターにおける検索時間の要求は7msec,30fps[22]の動画処理では33.3msecである。その場合、推論のためにデータをクラウド等へアップロードすることは時間的遅延が大きくなってしまうため、エッジ、あるいはフォグでの計算処理が必要となる。また、データ送受信の通信コスト、信頼性の観点からも、エッジにおける処理が望ましい。

1.7.4.2 組み込み型プロセッサ：推論の高速化・省エネルギー化

現在、ディープラーニングが活躍しているエンドユーザアプリケーションは、多くの場合、推論も学習もデータセンターに存在するバックヤード系の大規模なクラウドやスーパーコンピュータで行うのが主流である。送付するデータ量が少なく、かつある程度の時間が許容できる場合は、セミリアルタイムのアプリケーションですら適用が可能で、代表的にはBaiduのスマートフォンでの「リアルタイム」の音声認識・自動翻訳などが挙げられる。

しかしながら、自動運転やスマートシティ、ロボット制御などのリアルタイム系のIoT系アプリケーションでは、クラウドにデータを送付することなく、学習済みのネットワークを用いた画像や音声の推論・認識をエッジのデバイス側で行うことが求められる。

DNNを用いた推論では、少ない消費電力で学習後のDNNのフィードフォワード計算が必要十分な精

※21
OECD “Data-Driven Innovation For Growth and Well-Being,” Interim Synthesis Report.

※22
frames per secondの略。

度で行えることが必要となる。このような推論はIoT、ロボット、自動運転車に代表されるハードウェア的に独立した場所での用途が想定されるため、スーパーコンピュータを用いた学習のように大電力を計算に用いることは容易ではない。よって、このような推論を行うために、5〜15Wといった組み込み系レベルの低消費電力でフィードフォワード計算を行う用途に最適化された専用チップが開発されている。これらのチップは電力を極限まで削減するために、学習よりも低い精度で計算を行う機能が実装されているものもある。推論は学習と違い反復的な計算ではなく、出力もクラス分類等の計算誤差が無視されやすいことが多いため、このような低精度での計算も依然実用的であると考えられている。

- NVIDIA DRIVE PX（Tegra）：
  「Tegra」とはNVIDIAが開発しているモバイル／IoT向けのプロセッサシリーズであり、ARMプロセッサとGPU等がSoC（System on Chip）として搭載され、様々な省電力化の機構が組み込まれている。NVIDIAは自動運転を想定したプラットフォームである「DRIVE PX」シリーズでTegraチップを採用している。最新の「NVIDIA DRIVE PX2」では、一体化されたSoCのTegra（Parker）チップの10W版から、最大4個のPascalアーキテクチャのGPUを搭載している高機能版まで、自動運転の度合いによって複数の種別のものが提供されており、複数のカメラやセンサからの入力を、最大24テラフロップスの性能のDNNを用いて処理できるとしている。
- FPGAによる実現：
  FPGA（Field Programmable Gate Array）は特定の計算を電子回路レベルで実装できるため、同等の計算をGPUで行う場合よりも低消費電力であるといわれる。またディープラーニングは積和演算が主である計算であるために、特に有効である。スヨグ・グプタ（Suyog Gupta）氏らは、固定小数点型を用いたGEMM演算に特化した演算器をXilinx Kintex325T FPGAを用いて実装した。少ないbit数の固定小数点型を用いたディープラーニングでの、過度なアンダーフローによる丸め誤差を防ぐため、確率的に丸め処理を行う機構の擬似乱数生成回路等を用いて導入した。ディープラーニングに必要な計算精度は未だ自明ではなく、このような従来の計算機では計算効率が低下する特殊な計算手法が必要となる場合がある。よってFPGAを用いて加速を行うことは、今後の組み込み型のディープラーニング用アクセラレータとして活躍する可能性がある。
- ディープラーニング用ASIC:
  GPUベースのSoCは汎用性が高く、様々なネットワークやアルゴリズムが容易に適用できるが、それらがある程度固定化した場合は、やはり専用のASIC（Application Specific Integrated Circuit）の電力効率はGPUを凌駕する。多くの研究やスタートアップなどが早くから組み込み専用のASICを提案している。一例としてティエンシー・チェン（Tianshi Chen）氏らはディープラーニングに必要な積和演算に特化したアクセラレータ「DianNao」を提案した。DianNaoは485 mWの消費電力で452 GOP/s（ただしOPは16-bit固定小数点数の積和回数）の演算が行え、1Wあたり1Topsと、同世代のTegra K1と比べて約10倍の電力効率を達成している。

全体的に俯瞰すると、IoT向けの推論が汎用性の高いGPU型になるのか、あるいはASICのように専用化するのか、あるいは中間的なものになるのかの決着は付いていない。IoTのエッジデバイスは大量生産される可能性が高いため、通常では他のIT分野同様にASIC化によるコストメリットや省電力化が非常に効果的なはずである。

しかしながら、現状では、後述のディープラーニング専用プロセッサ同様、学習アルゴリズムの高度化・高速化の発展が非常に目覚しく、特定のASICを開発してもそれらの進化に追随できず、早期に陳腐化してしまう可能性も高い。よって、しばらくは主に上記三つのプラットフォームが切磋琢磨する状

況が続くと予想される。

1.7.5 エッジ、フォグ、クラウドの役割の最適化

学習と推論の両方を実施する場合には、その双方を考えて、エッジ、フォグ、クラウドのどこで、何を、どの程度の頻度で計算するかについて、最適化が必要となる。基本的には、多くの末端のデバイスからのデータをなるべく統合して学習を行ったほうが精度向上のスピードアップが期待できる。そのため、推論の計算処理を随時行いつつ、学習のためのデータはエッジから見て上層（フォグ、クラウド）にアップロードし、上層で学習を行った後に、学習後のモデルをエッジにダウンロードして推論に利用することが考えられる。

ただし、全データを上層にアップロードするのでは結局トラフィックの問題が生じる。そのためには、学習に利用すべきデータとそうでないデータを峻別する技術が必要となるだろう。その最適化された姿は応用分野に依存するが、例えば、推論精度が悪いデータのみを峻別して学習用データとしてアップロードすること等が仕組みとして考えられる。

上層におけるクラウド側には、学習済みモデルのレポジトリと、モデルのアップデートを行う機能が求められる。民間のクラウドサービスでも、学習のためのインフラ提供機能だけでなく、上記の様な機能も合わせて提供する会社は増えるであろう。

また、エッジやフォグで利用される推論用の計算デバイスには、学習用の計算デバイスに比較して、低消費電力であることが求められる場合が多い。したがって、消費電力の大きいGPUではなく、搭載する計算ロジックを書き換えることができるFPGAや専用チップのニーズが高まると考えられる。Intelは、FPGAを主力製品とするAlteraを買収し、CPUとFPGAを同一パッケージや同一ダイに統合した製品の開発を行っており、Xilinxは画像処理に特化した開発環境reVisionを発表している。これらはFPGAで推論に特化した回路を想定しており、学習までには対応していない。同様なツールはサードパーティーからも発表・リリース予定があり、例えばTERADEEP（米国）、DeePhi Tech（中国）は既にFPGAを採用した推論アクセラレータをリリースしている。日本でも同様のリリースが行われるであろう。

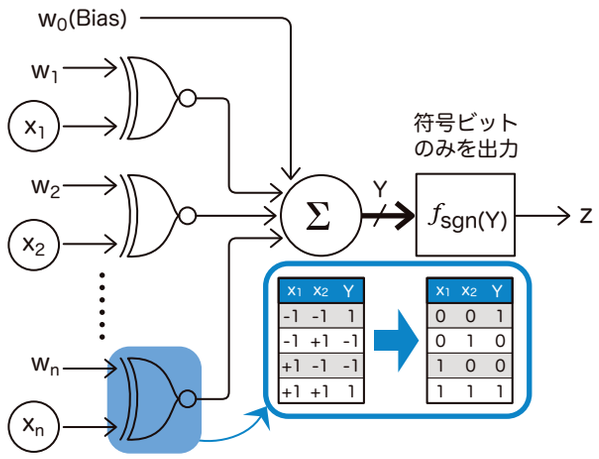
ディープラーニングの学習や推論において、必ずしも倍精度計算は必要なく、8ビット、4ビット、更には1ビット（2値化）／2ビット（3値化）での計算も可能との研究結果が相次いでおり、消費電力や計算性能の観点からFPGAを用いたプロトタイプの研究開発が進んでいる。近年の研究開発動向を**表6**に示す。興味深いのはIntel、Xilinxがともに1ビット／2ビットの研究開発を進めていることであり、近い将来これらの精度が実用化される可能性は高い。

■表6　FPGAにおける2値化/3値化DNNの開発動向[23, 24]

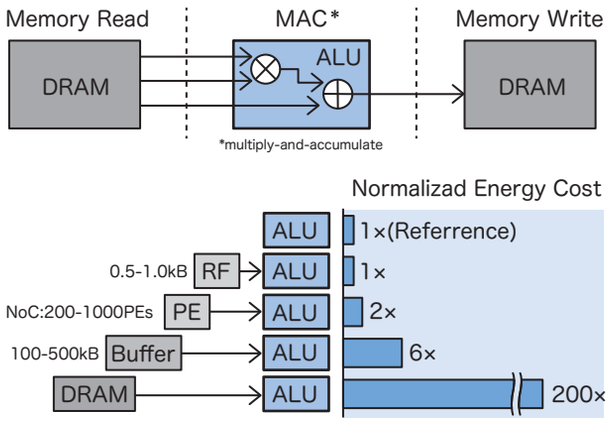
FPT2016（2016 年 12 月開催）
E. Nurvitadhi（Intel）et al., “Accelerating Binarized Neural Networks: Comparison of FPGA, CPU, GPU, and ASIC”
H. Nakahara et al., “A Memory-Based Realization of a Binarized Deep Convolutional Neural Network”
ISFPGA2017（2017 年 2 月開催）
Ritchie Zhao et al., “Accelerating Binarized Convolutional Neural Networks with Software-Programmable FPGAs”
Y. Umuroglu（Xilinx）et al., “FINN: A Framework for Fast, Scalable Binarized Neural Network Inference”
H. Nakahara, H. Yonekawa, “A Batch Normalization Free Binarized Convolutional Deep Neural Network on an FPGA”
Y. Li et al., “A 7.663-TOPS 8.2-W Energy-efficient FPGA Accelerator for Binary Convolutional Neural Networks”
G. Lemieux, “TinBiNN: Tiny Binarized Neural Network Overlay in Less Than 5,000 4-LUTs”

※23
“The 2016 International Conference on Field-Programmable Technology (FPT ’16).” ICFPT2016 Website
<http://www.icfpt2016.org/>

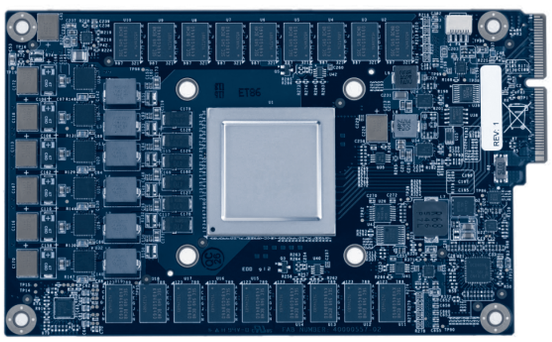
※24
“25th ACM/SIGDA International Symposium on Field-Programmable Gate Arrays.” FPGA2017 Website
<http://isfpga.org/>



■図38　2値化ニューラルネットワークの概要



■図39　演算器と各種メモリの電力コスト[25]



■図40　TPU搭載ボード[26]

■図41　TPU搭載サーバ[27]

図38に2値化ニューラルネットワークの概要を示す。通常、GPUやCPUではニューラルネットワークの基本演算である積和演算を8ビットや16ビットの精度で行う。2値化、すなわち1ビット精度でこれを行うと、最も電力・面積を必要とする乗算回路をXNORゲートで実現できる。したがって、大量の積和演算回路を1チップ上に集積することができる。もう一つの利点は、ニューラルネットワークの重みと計算結果を格納するバッファ（メモリ）のサイズを大幅に小さくできることである。ニューラルネットワークの構成によっては全てのバッファを1チップに格納できるので、DRAM（Dynamic Random Access Memory）のようなオフチップは不要になる。

図39に演算器と各種メモリの電力コストを示す。積和演算回路（Arithmetic Logic Unit; ALU）と各種メモリの距離に電力は比例する。2値化により全てをオンチップに格納できればDRAMに要する電力をバッファやレジスタファイル（RF）で必要な電力に抑えることができる。したがって、2値化は電力削減にも貢献する。また、オンチップメモリで格納できれば、演算器チップとオフチップの通信ボトルネックも解消できる。

※25
Massachusetts Institute of Technology, Energy-Efficient Multimedia Systems Group, The Eyeriss Project, “Tutorial on Hardware Architectures for Deep Neural Networks”

※26
「GoogleのTensor Processing Unit (TPU)で機械学習が30倍速くなるメカニズム」Google Cloud Platform Japan Blog Website <https://cloudplatform-jp.googleblog.com/2017/05/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu.html>

※27
同上。

したがって、低精度によるオンチップ化は今後ますます研究開発が進むとみられている。画像の物体認識を行う際、GPU（NVIDIA Tesla K40）では235Wかかるところ、2値化を導入したFPGAでは4.7Wで可能になる。また、モバイルGPU（NVIDIA Jetson TK1）と比較して2値化を導入したFPGAは15倍高速に処理できることが報告されている[7]。

学習時と推論時で必ずしも同じデバイスを利用する必要はなく、むしろそれぞれの用途に特化したデバイスとすることにより高い計算性能や低い消費電力を実現する試みもある。Googleはディープラーニング用チップであるTPU（1.7.1項、**図40**、**図41**参照）を開発し、既に利用を開始していることを2016年5月に発表した。消費電力を従来のデバイスに比べて10分の1程度に抑えられたとともに、Googleのディープラーニング向けオープンツールであるTensorFlowとの親和性を高め、開発者はハードウェアの違いを意識せずに利用できるようになっているとしている。また、韓国ではKAISTの（イチョン・ユー）氏Hoi-Jun Yooらのグループが、CMOS（Complementary MOS）を用いたディープラーニングの推論用のデバイスの開発を行っている。

画像処理などディープラーニングの適用が進んでいる一部の分野については、エッジ側での推論に係るハードウェアとソフトウェアの構成に関して実装例が提案されつつある。実装の方向性として、現状のCNNの枠組みを前提としてハードウェアを最適化してしまう方向と、専用化する部分とGPU、FPGAを一部利用して柔軟性を残しておく方向がある。現状のディープラーニングの研究が日進月歩である状況を考えると、研究の現状を前提とするよりも、今後の発展部分を採り入れる柔軟性を残しておく後者は、最高性能は劣る可能性があるものの、適用分野の幅広さや息の長い開発を見込むことができるだろう。

1.7.6 次世代AIインフラストラクチャ・ハードウェア

ディープラーニングによるAIが大きな成功を収めつつあることを背景として、ディープラーニングに向いたアーキテクチャを備えたプロセッサ、あるいは脳型コンピュータと呼ばれるアーキテクチャを構想し、実装しようとする動きが各所で起こっている。この項では、主に、脳を参考にしたモデル[28]に基づいて構築されるディープラーニングのためのニューラルネット計算向きアーキテクチャについて概観する。

1.7.6.1 AI向き専用マシンの歴史

コンピュータは、プログラムの入れ替えによってどんな問題にも対処できる万能マシンとしての性能を価値としてきた。しかし、汎用機は大型で高価であることから、特定の問題向きの専用機であれば、より小型で安価な解があるはずだと考えられることはしばしばあった。AIが米国から世界に広がり始めた頃は、AIプログラムとは、Lispなどの記号処理に適した言語で書かれたプログラムのことであり、記号処理に向いたアーキテクチャが構想された。1980年頃には、Lispマシンが商用化された。Lispマシンは、タグ付けされたデータとメモリ管理に適した構造を持つが、その内容は、ノイマン型計算機であり、記号処理向きの命令をマイクロコードで実現していた。

1982～1992年に実施された日本の第5世代コンピュータもAI向きのコンピュータを目指した。AIの本質は論理であるとの考えから、論理型言語Prolog向きのコンピュータが実装されたが、内容はやはりノイマン型計算機にProlog向きのプリミティブをマイクロコードで実装していた。1990年頃からの誤差逆伝播学習の発明に基づく第2期AIブームでは、当時流行していた計算プラットフォームであるワー

※28
Hodgikin-Huxleyのパルス神経伝達モデル、Hebbの学習則、McCulloch-Pittsのニューロンモデル、Hubel-Wieselの神経階層仮説などがある。

クステーションに増設ボードとして付加できるニューラルネットワーク計算のアクセラレータが発売されたが、大きな勢力となることはなかった。

汎用計算機に対抗して開発されたAI専用のコンピュータであるが、AIに限らず、専用計算機が成功するのはまれである。上記のAIコンピュータの中では、Lispマシンは数千台販売されたが、ほかは大きなマーケットを形成していない。大きなマーケットを占めることができた専用プロセッサは、GPGPUだけであろう。専用プロセッサは、その名のとおり、応用領域を限定しているため、汎用計算機に比べて小さな市場にとどまるのは当然である。後述するように、GPGPUはAI向けにも重要な地位を占めているが、GPGPUはその名のとおり、GPUという専用プロセッサをGP=general purposeに適用できるようにしたために利用が進んだ。同様に、FPGAも汎用性の高い専用プロセッサの基材であるために一定の市場を持つことができたといえる。

このように考えると、今盛んに行われているAI向けのニューラルネットワークプロセッサが、成功を収める可能性については未知数である。しかし、今回のAIプロセッサブームは、以前のAIコンピュータブームとは異なる面がいくつもあるので、大成功する可能性も秘めている。それについては後述することにして、まず、現在のニューラルネットワークプロセッサの実例を概観する。

1.7.6.2　AIプロセッサの分類

従来型のコンピュータに対するAIコンピュータ、脳型コンピュータの位置付けをノイマン型、非ノイマン型コンピュータの区別とともに**図42**に示す。AIの情報処理に向いたコンピュータをAIコンピュータとして黒枠で示している。この中には、ニューラルネットワークだけでなく、例えば論理計算に向いたコンピュータも含まれる。現代の計算機を構成する電子回路を論理回路と呼ぶことからも分かるように、ノイマン型コンピュータは、論理計算の機能があり、したがって、前述の第5世代コンピュータもノイマン型のAIコンピュータと位置付けられる。

現在注目を浴びているAIコンピュータは、青枠で示したニューラルネットワーク型、あるいはニューロモーフィック型である。現代のコンピュータは、計算論的に万能マシンであるので、ディープラーニング計算をPCやスーパーコンピュータ、あるいはGPGPUで実行させることが可能である。特にGPGPUは、数千を超えるコアを備えて高い並列性を発揮できるので、ノイマン型のなかでも非ノイマン型に近いアーキテクチャを持つ。

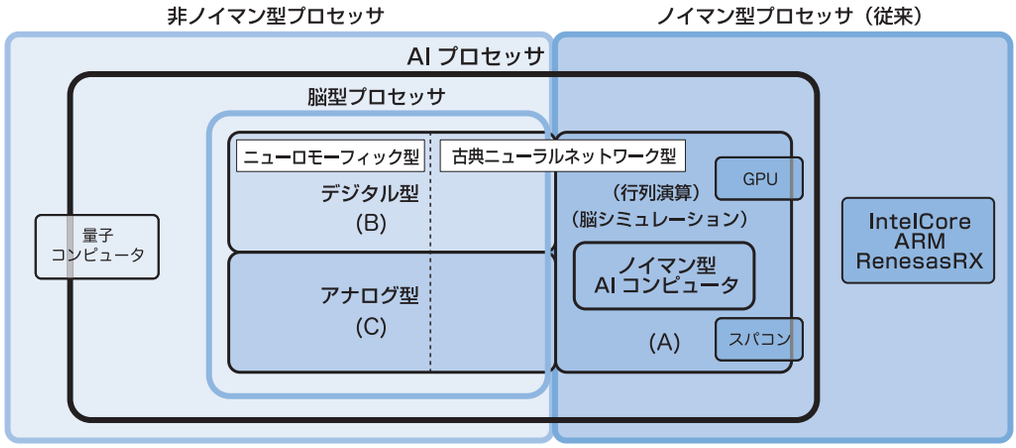
脳型コンピュータでは、シナプス結合の強度が、記憶の役割を果たす。その記憶の実現方法には、デジタルコンピュータと同様にDRAM等のメモリを用いるデジタル型と、メモリスタあるいはReRAM[29]のような電気抵抗変化を記憶できるデバイスを用いるアナログ型がある。両者のシナプスとニューロンの細胞体に対応するアナログとデジタルでの実現回路例を図43に示す。

1.7.6.3　スーパーコンピュータとGPUによるニューラルネットワーク計算

本格的なニューラルネットワークプロセッサの前に、スーパーコンピュータとGPGPUによるニューラルネットワークのシミュレーションについて述べる。スーパーコンピュータとGPGPUは、PCやウェブサーバなどから見ると特殊なハードウェアであるが、ソフトウェアによってニューラルネットワークシミュレーションを行う点では、PCによる実行と違いがない。現在、機械翻訳やゲームなどで高性能を発揮しているのは、これら通常型のAIプロセッサである。

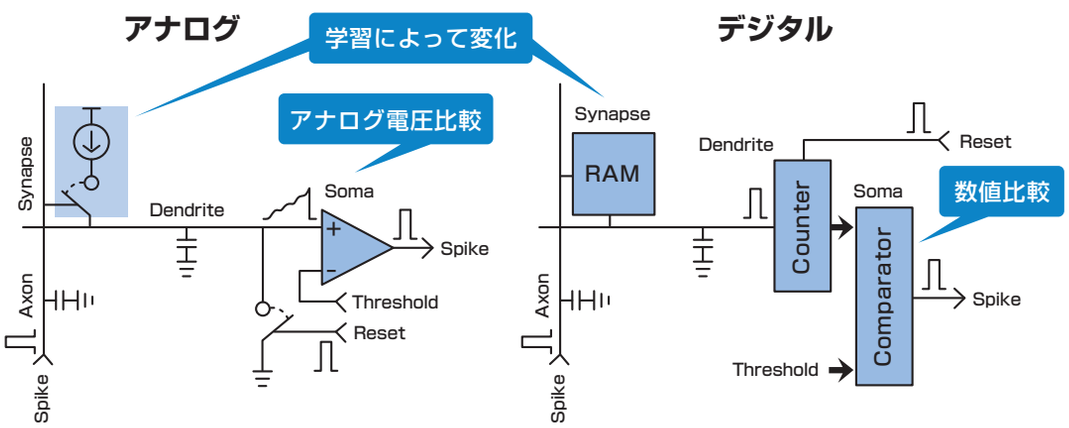
ニューラルネットワークのシミュレーションには、**表7**のような種類のソフトウェアがよく使用され

※29
抵抗変化型メモリ(Resistive Random Access Memory)。電気抵抗の変化を利用したメモリ。



【語句定義】
ノイマン型プロセッサ：一つのメモリにデータとプログラムを内蔵、メモリから命令を逐次取り出しプロセッサで実行
非ノイマン型プロセッサ：ノイマン型以外
AI プロセッサ：機械学習、深層学習の演算処理を行うハード（プロセッサ、メモリ等の集合体）
ノイマン型 AI プロセッサ：【図中（A）】
脳型プロセッサ：AI プロセッサの中で、ニューロン・シナプスのような脳機能を使った演算処理を行う
デジタル型脳型プロセッサ：ニューロン機能をデジタル素子で模擬する脳型プロセッサ【図中（B）】
アナログ型脳型プロセッサ：ニューロン機能をアナログ素子で模擬する脳型プロセッサ【図中（C）】

■図42　AIコンピュータ、脳型コンピュータの位置付け[30]



■図43　シナプスとニューロン細胞体の実現回路例[31]

ている。これらの多くはオープンソースである。ソフトウェアとしての実績が蓄積すると、それをLSI化して高速化しようとする動きが出てくる。例えば、GoogleのTensorFlowは、ニューラルネットワークソフトウェアとしても定評があるが、Googleは、TPU（TensorFlow Processing Unit）としてハードウェア化を図っている。テンソル（tensor）というベクタ、マトリクスなどのデータの並べ方を一般化した数学的概念をその名に冠していることから、ニューラルネットワークに限らず非常に汎用的な数学ライブラリをベースにしていることをうかがわせる。

日本では、スーパーコンピュータ「京」を保有する理化学研究所が、2013年8月に、10兆個の結合を持つ神経回路のシミュレーションに成功したことを発表している[32]。ニューロン数は17.3億個であり、小型霊長類の全脳の規模に達している。

※30
NEDO技術戦略センター作成
※31
文献[8]より作成。
<http://www.riken.jp/pr/topics/2013/20130802_2/>

※32
「「京（けい）」を使い10兆個の結合の神経回路のシミュレーションに成功－世界最大の脳神経シミュレーション－」『広報活動』理化学研究所ウェブサイト

■表7　ニューラルネットワークのシミュレーションソフトウェア[34]

名称	開発者、頒布	言語、プラットフォーム	普及状況
Tensorflow	Google（米） オープンソース	Linux、MacOS、GPU 可 Python、C++	ベンチマークも含め、標準 Google 翻訳などにも使われる
Torch	DeepMind（Google） オープンソース	Lua（言語）、Google Cloud Platform	AlphaGO が使用、並列化が強力
Chainer	PFN（日） オープンソース	Linux、GPU 可、 Python	コミュニティが、ほぼ国内のみ。 画像処理から自然言語処理、ロボット制御 （NEDO「次世代人工知能ロボット中核技術開発事業」で拡張中）
Caffe	UC Berkeley（米） オープンソース	Windows、MacOS、 Linux、GPU 可、C++、 （Python）	最も古く、コミュニティが大きく、 サポート OS が多い。画像認識専用で高速
PyLearn2 （Theano）	Montreal 大（加） オープンソース	Windows、MacOS、 Linux、GPU 可、 （Python）	Caffe に次いで歴史がある
Neural Network Toolbox	Mathworks（米） 商用	Matlab の一部、GPU 可	制御システムや組込みソフトウェアの 開発ツールである Matlab のモジュール と組み合わせられる

■表8　ニューラルネットワーク計算のためのプロセッサチップ（ノイマン型）[*34]

開発者	国	名称	アーキテクチャ	実績・特徴	出荷時期
NVIDIA	米国	Pascal	153 億個トランジスタ、16nm FinFET 16GB の DRAM を 720GB/s 接続	AI 研究のスタンダード、産業技術総合研究所 ABCI にも採用予定	出荷済み
Google	米国	TPU Tensor Processing Unit	Google 製 NN シミュレータ TensorFlow のハードウェア化	低精度の浮動小数点演算で高速化 Google 翻訳、AlphaGo、ストリートビューやクラウドサービスで採用中	外販しない
Microsoft	米国	Catapult	17 万の ALM と、1,600 の DSP を集積した ALTERA 製 FPGA	Bing サーチに使用	未公表
デジタルメディア プロフェッショナル 産業技術総合研究所[*36]	日本	EP1	GPU のシェーダプロセッサに DNN 高効率処理 HW を接続	1W 電力により、広く組み込みシステムでの適用を目指す	2025 年
富士通	日本	DLU	京の成果の Tofu インターコネクトで 10 万チップの接続	未公表	2018 年
Deep Insights （PEZY グループ）	日本	DI-1	7nm プロセス、100 万コア	チップ間の磁界結合、ノード間の Si フォトニクス技術を用いた光通信 100TB/ 秒、３D 積層 DRAM、液浸冷却などの実装技術	2018 年
理化学研究所[*37]	日本	未公表	低精度計算と高並列化、NVIDIA T100 の 70 倍が目標、40nm TSMC で製造	半精度化により倍精度の 16 倍を達成する	2017 年度チップ試作
Intel（Nervana）	米国	Lake Crest	HBM2 独自のインターコネクト、1TB 秒	深層学習フレームワーク “Neon” 用意	2017 年　特定顧客向け出荷
Graphcore	英国 韓国	IPU	未公表	DNN モデルをプロセッサ内部に保存	2017 年　PCI カード出荷

GPUは、もともとグラフィックス計算のためのプロセッサであった。画面中のたくさんの頂点に対する座標変換や、ポリゴンへのテクスチャマッピングを高速処理するためにSIMD並列計算する機能を

※33
MMX（Multimedia Extensions）、SSE（Streaming SIMD Extensions）はIntelが開発したCPUの命令セット。
※34
各種公表資料よりNEDO技術戦略研究センター作成。
※35
NEDO「戦略的エネルギー技術革新プログラム」でPEZYグループのPEZY Computing（日本）が開発し、理化学研究所に設置した「Shoubu」（菖蒲）が、2016年にスーパーコンピュータのエネルギー効率ランキングである「Green500」で3期連続の世界1位を獲得した。

備えている。このSIMD計算のことを「NVIDIA」は、Single Instruction Multiple Threadと呼んで、インテルアーキテクチャのMMX、SSE[33]と区別している。GPUの並列計算機能をグラフィックス以外にも応用するために、NVIDIAは、CUDAというC言語を拡張したプログラミング言語、プログラミング環境を提供し、General Purpose GPU、すなわちGPGPUとしての計算モデルを構築した（**表8**）。

GPGPUは、頂点の座標変換のための積和演算が得意な構造をしている。すなわち、条件分岐が混じらない単精度浮動小数点演算でよい性能を上げる。単精度の積和演算は、ニューラルネットワークの主要な演算であるため、ニューラルネットワークシミュレーションの高速化にも大きな効果を発揮する。x86などのPCやエンタープライズ系CPUでは、2000年頃まで、単精度より遅い倍精度浮動小数点演算を単精度と同じ計算速度にすることに努力が払われたが、グラフィックス計算では、たかだか12～16ビットの計算精度で足りるので、GPUは、32ビット単精度どころか、16ビットの半精度の計算速度を上げることに注力されている。ニューラルネットワークの順方向計算には、16ビットより更に短い8ビット、極端な場合は、1ビットでよいとする説がある。この計算精度の節約は、並列度の引き上げと省電力性の改善のためのキーテクノロジである。

日本でも理化学研究所、PEZYグループのDeep Insightsなど、複数の機関が短精度化を絡めたプロセッサ開発を進めている[35]。短精度計算は、順方向計算には適用できるが、学習のためには、係数の勾配を計算できなければならないので、精度を落とせないことに注意が必要である。

表7に掲げたニューラルネットワークソフトウェアは、ほとんどが、CPUだけでなくGPGPUで実行するコードを生成する。GPGPUは、ニューラルネットワークの重要な計算プラットフォームとなっている。

1.7.6.4　デジタル型のニューラルネットワークプロセッサ

スーパーコンピュータやGPGPUによるニューラルネットワークは、汎用のコンピュータをニューラルネットワーク計算にも応用しようとする試みであるが、本項で述べるのは、ニューラルネットワーク計算のために設計、試作されたプロセッサチップである（**表9**）。

■**表9**　ニューラルネットワーク計算のためのプロセッサチップ（非ノイマン型）[*34]

開発者	国	名称	アーキテクチャ	実績・特徴	出荷時期
IBM	米国	TrueNorth	データフロー型ニューロン 100 万個、シナプス結合 2.56 億個、54 億トランジスタ、SRAM	順方向認識処理のみチップ当たり 70 m W、16 チップ接続可能	2014 年発表
Wave Computing	米国	DPU	独自のデータフロー型、16nm FinFET、16,000 コア、32GB、512GB DDR4	TensorFlow を初期状態でサポート	2017 年 4Q 量産
東芝	日本	TDNN	神経細胞の興奮、結合係数を 1 ビットに圧縮し、32 k シナプスを 1.9㎜角に集積	興奮レベルをゲート遅延で表現し、可さんする信号処理で極端な省電力化	2016 年 11 月学会発表
トプスシステムズ [*38] NEDO 次世代人工知能	日本	SMYLEdeep	データフロー型　8 コア 75 M H z	低い動作周波数で消費電力を抑えつつ、最大 480fps で超高速画像認識処理が可能	2017 年 2 月 プレスリリース

表9の中では、1.7.1項で述べたIBMの「TrueNorth」が傑出している。4,096個のプロセシングエレメントによって100万個のニューロン、2.56億個のシナプス結合の並列処理を実現し、28nmプロセス[34]

※36
NEDO「IoT推進のための横断技術開発プロジェクト」で開発中。

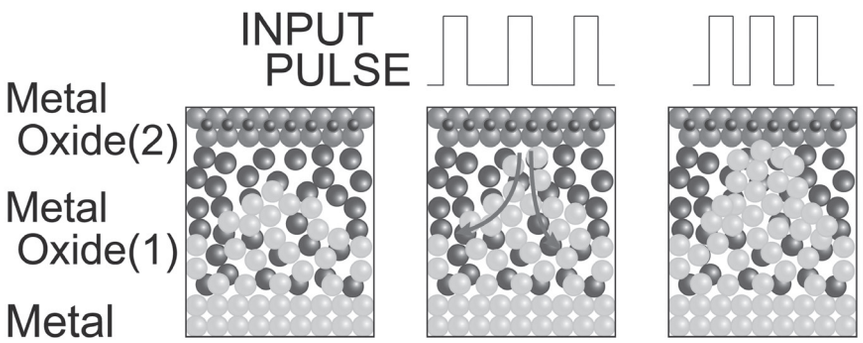
※37
NEDO「次世代人工知能・ロボット中核技術開発」事業で開発中。

※38
NEDO「IoT推進のための横断技術開発プロジェクト」で開発中。

で１チップに集積した。16個を組み合わせることが可能である。省電力性能も高く、400×240のビデオを30fpsで認識するのに要する電力は63mWに過ぎない。

1.7.6.5 アナログ型のニューラルネットワークプロセッサ

デジタル型ニューラルネットワークでは、シナプスの結合強度の表現と記憶にDRAM等のデジタルメモリを使用するが、アナログ型ではアナログメモリを使用する。1960年代、アナログコンピュータがデジタルコンピュータに駆逐された主たる理由は、アナログ電圧を記憶するメモリが作れなかったことであるが、2008年にHPが「メモリスタ」（memoristor）を発見すると、通過する電荷量に応じて変化する電気抵抗値をメモリに応用する研究が行われた[39]。



■図44　電気抵抗メモリの変化[40]

この電気抵抗メモリは、特に不揮発性に注目されたが、ニューラルネットワークへの応用では、電気抵抗値が連続に変化することに価値がある。産業技術総合研究所で研究が行われているRAND（Resistive Analog Neuro Device）では、電荷が流れることで、物質内に**図44**に示すような物理的変化が引き起こされ、電気抵抗が変化するとされている。

フラッシュメモリがマルチレベル化で集積度を上げたように、アナログ方式では、一つのメモリ素子にデジタルでいう多ビットを重畳できるので、小型化に適していると考えられる。しかし、アナログ回路は、デジタル回路のように極端な微細化に適していない。デジタル回路は、閾値に幅を持たせてノイズやばらつきを許容することができるが、アナログでは、広い動作範囲での応答の線形性やノイズ耐性を求められるからである。

表10に掲げたアナログ型ニューラルネットワークプロセッサは、シナプス結合強度の表現に、次世代メモリとして研究されていた不揮発メモリ素子を活用している。これらのメモリは、長期間研究が続けられているが、対抗する従来型のメモリであるDRAMとNANDフラッシュメモリも高集積化が続いているので、世代交代を果たせていない。次世代メモリに関する企業や研究者が、ニューラルネットワークプロセッサという新たな可能性を見出して殺到していると解釈できる。

この中で、東芝のTDNN（Time Domain Neural Network）は、シナプス結合強度をバイナリで表現するため、デジタルとアナログの違いよりも、係数を電圧ではなく時間遅れで表現する点に特色がある。これによって従来の試作品の6分の1の消費電力を達成している。一般に、アナログは、デジタルよりも大きな省電力効果が得られる可能性がある。

※39
DRAMは、コンデンサに蓄えられる電荷、フラッシュはフローティングゲート中の電荷、MRAMや磁気ディスクは磁気によって記憶する。ReRAM、PcRAM、RRAMなどの次世代メモリは、電気抵抗の変化を記録する。STT-MRAMもスピンによって生じる電気抵抗の変化を取り出す。

※40
「Aidevice」semiconportalウェブサイト
<http://www.semiconportal.com/aist_aidevice/>;
「Aidevice Artificial Neural Network System for Inference」
Aidevice ウェブサイト <http://green-innovation.jp/aidevice/>

■表10　アナログ型ニューラルネットワークプロセッサ [41]

開発者	国	発表時期	シナプス可塑性の実現素子	概要
IBM	米国	2016 年 5 月	PCM（GeSbTe）	シナプスを PCM で構成、integrate-and-fire 型ニューロンの挙動を再現 1 チップ上に 400 万セルを集積化 AI チップと組み込みプロセッサ月の FPGA と実
パナソニック	日本	2013 年 6 月	FeMEM	CMOS 型ニューロンの配線層にシナプス
Posteck SK Hynix 等	韓国	2015 年 12 月	メモリスタ（TiN/PCMO）	ニューロンとシナプスを別々のチップに実装
デンソー University of California, Santa Barbara	米国 日本	2015 年 12 月	メモリスタ（Al203/Tio2）	CMOS ニューロンの上に 12x12 クロスバー構造のシナプスを形成 2030 年　車載用実用化目標
HP、 University of Utah、 University of Michigan	米国	2016 年 12 月	メモリスタ	32nm プロセスを用いた場合 チップ面積 85.4mm^2 メモリスタを用いた画像認識への取組み
東北大学	日本	2016 年 12 月	スピントロニクス素子	スピントロにクス素子 36 個と FPGA との組み合わせ
産業技術総合研究所　他	日本	2016 年 6 月開発開始	アナログ型抵抗変化素子	対 TrueNorth 電力効率 100 倍、チップ面積 1/20 28nm プロセス、100 万個以上のシナプスを集積化する技術を開発
NEC 東京大学	日本	2016 年 9 月共同開発発表	未公表	ブレインモルフィック AI 技術 東京大学合原教授が中核
東芝	日本	2016 年 11 月に学会発表	SRAM/ReRAM	時間領域アナログ信号処理技術による小型 / 省電力化、消費電力 1/6

1.7.6.6　次世代ニューラルネットワークプロセッサの方向性

ニューラルネットワーク及びより生物脳に忠実なニューロモーフィックコンピューティングは、専用マシンによらずとも、万能マシンたる現代のコンピュータで実行することが可能である。実際、スーパーコンピューティングにより世界最大規模の脳シミュレーションが実施されたほか、GPGPUには、多数の応用がある。

コンピュータは、プログラムを入れ替えることでどんな情報処理にも適用できる汎用性に大きな強みがあり、応用先を限った専用機は、必ず苦戦することは先に記述した。それにもかかわらず、ニューラルネットワークプロセッサを開発する価値は、ノイマン型アーキテクチャからの脱却と、大規模なニューラルネットワークを構築することにある。半導体産業に陰りの見えた日本にとって、新たなチャンスとなるかもしれない。

ノイマン型アーキテクチャとは、プログラムとデータを同じメモリに内蔵すること、メモリからの命令に従って演算器を逐次的に動作させることでハードウェア量を減らすこと、プログラムをデータのように操作することができるコンピュータであり[42]、現在主流となっている。このアーキテクチャでは、メモリからのプログラム（命令）の読み出しとデータの読み書きが輻輳し、フォン・ノイマン・ボトルネックと呼ばれる情報伝達の隘路が生じることが問題視されてきた。

※41
各種公表資料よりNEDO技術戦略研究センター作成。

※42
プログラム内蔵型のアイディアを出したのは、フォンノイマンではなく、ENIACを開発したエッカートとモークリーである。フォンノイマンは、1946年頃にこのアイディアを聞き及び、軍事機密の禁を破って、フォンノイマンの単独名でレポートを作成した。ENIAC開発予算を工面したゴールドスタイン中尉と成果を宣伝したいペンシルバニア大学は、このレポートを広く公表してしまった。これによってフォンノイマンの名前が知れ渡ったが、エッカートとモークリーはアイディアを公知にされたおかげでUNIVACから獲得していたコンピュータ特許を無効とされた。

その解決は、多数の演算器を用意し、その近傍に専用のメモリを置く方法が考えられるが、ニューラルネットワークアーキテクチャは、まさにそのような構造になっている。現在のコンピュータは、半導体のムーアの法則とスケーリング則を信奉して50年の進歩を続けてきたが、微細化の限界が近づいており、両法則の推進だけでは発展が望めなくなっている。ニューラルネットワーク型アーキテクチャによって、新たな発展軌道に乗れるのではないかとの期待がある。

大規模なニューラルネットワークを目指す動きは、人間の脳のニューロン数（大脳皮質で140億個程度、小脳など全ての脳細胞を足すと1,000億個近いとされる）に大きな意味があるのだとする立場から生じている。チンパンジー等、高等な霊長類の大脳皮質のニューロン数は、60～80億個である。チンパンジーは、高度な知的能力を発揮するが、人間が、言語や記憶を駆使して文明社会を形成する能力とは雲泥の差がある。

SRI Internationalのハンス・モラベック（Hans Moravec）氏が2000年頃に発表した計算機の発展と生命体の知能の比較によれば、現在の人工脳の到達レベルは、ネズミのレベルに相当する。その処理能力をあと3桁ほど伸ばすことで、人間レベルに達するのであれば、その努力を続ける価値があろう。いろいろな生物脳のニューロン単体の能力はほとんど一定であると推測されるので、処理能力の差は、ほぼニューロン数とシナプス結合数の違いから生じていると推定される。脳細胞数を増やす重要なアプローチは微細化であるが、前述したようにムーアの法則には陰りが見えていると同時に、消費電力の増大が問題となる。

NVIDIAの新しいGPGPUである「Pascal」は、170 テラフロップスという初代の地球シミュレータ[43]を上回る性能を示すが、300Wを消費する。人間の脳は、消費エネルギー20Wであり、大きな差がある。その原因は、PCのCPUに比べれば一つ一つのプロセッサの性能を落としたGPGPUであっても、まだ無駄な計算能力を発揮していることが考えられる。

既に、バイナリ（2値）のニューラルネットワークであっても、ニューロン数をそれに応じて増やせば、より高精度のニューラルネットワークと遜色ない性能を出せることが示されている。IBMの「TrueNorth」は、演算精度と速度を落とすことで、100万ニューロンで70mWという高い性能／電力比を示している。すなわち、人間の脳に近いレベルのニューロン数をリーズナブルなエネルギー性能で実現するには、現在のコンピュータハードウェアは不適当であり、より生物の脳に近い脳型コンピュータに勝機があると推測できる。

現在のニューラルネットワークプロセッサ、あるいはニューロモーフィックプロセッサは、順方向計算の加速には大きな効果を発揮するが、学習能力は持っていない。計算量としては、誤差逆伝播学習は、順方向計算の数倍程度であるが、繰り返し回数が多いため高速性が求められる。また、誤差逆伝播学習では、ニューロンの興奮度から結合係数の変化量を決める際に係数の微分値が必要になり、微分値の計算にある程度の計算精度が必要となる。

順方向計算のように1～16ビットの係数精度では不足するので、現在は、GPU以上のプロセッサでないと実装が難しい。そのため、学習は、データセンターなどのサーバに任せ、エッジ側のプロセッサでは順方向計算に割り切るという分担ができている。しかし、人間が作業しながら学習するように、今後はAIも順方向計算と逆伝播学習を同時に並行して進めたいという要求は出てくるであろう。学習機能をハードウェア化できれば、リアルタイム学習のような道も開けるであろう。

ニューラルネットワークプロセッサの実現には、デジタル型とアナログ型があることを述べた。

※43
海洋研究開発機構（Japan Agency for Marine-Earth Science and Technology; JAMSTEC）に設置されているスーパーコンピュータ。
2002年に35.86TFlopsで世界1位の性能を記録した。

1970年頃まで、コンピュータにはアナログ方式とデジタル方式があったが、アナログ方式が敗退したのは、誤差が大きく、数値をメモリできないこと、メモリできないからコピーもできず、再現性が低くなることが原因であった。対するデジタル方式は、ワード長を変えれば精度をいくらでも上げることができ、閾値を設けて値を0、1に分離することにより、ノイズ耐性が向上し、したがって数値の長期記憶が可能となり、記憶をコピーして再現することができた。

ほぼ同じ議論が、アナログ式のニューラルネットワークプロセッサにも成り立つ。シナプス結合強度をアナログの電気抵抗値で表現する方法は、スペース効率と電力効率に優れるが、ノイズによって精度が劣化しやすく、情報を別のプロセッサにコピーして再現することが難しい。コピーするためには、電気抵抗値を計測して送り出す機能が必要となるが、そのための信号線を別途設けることにも困難がある。クラウドで学習した係数データを、多数の末端のアナログAIプロセッサに高速で焼き付けるのも難しい。

コピーが難しく再現性が低いということは、各々のAIプロセッサが個性を持つことを意味する。人間世界では、個性は重要なファクターであるが、機械の個性は必ずしも歓迎されないだろう。例えば、AIが事故を起こした場合に原因をソースプログラムに戻って検証することが困難であり、多数のデバイスに学習データをダウンロードして一律の改善を図ることも困難になる可能性がある。したがって、アナログ方式に特別の意味が見出されるか、アナログ値のコピーの方法が確立するまでは、デジタル式のAIプロセッサが作られ、使用されていくだろう。

1.7.6.7　脳に忠実なモデルと工学的に単純化したモデルとのバランス

脳型コンピュータは、人を超える知的能力を目標に発達していくと想定されるが、現在行われている単純なニューロンモデルでよいのかどうかは議論が分かれる。確かに、ディープラーニングは、特にパターン認識を中心に人間を超える計算能力を発揮しているが、言語、計画、論理、創造など、重要な知的能力の実現は発展途上である。ニューラルネットワークの学習に、大量のデータを何万回も繰り返し与えなければならない点も、生物の学習とは大きく異なる。したがって、ニューラルネットワークの更なる多層化や、ニューロン数を増やすだけで広範な人間的知能を獲得できるとは思われない。

このようなニューラルネットワークの限界を克服すべく、シナプス結合強度を単に積和で計算するのではなく、脳の中の信号の同期性や揺らぎに注目するニューロモーフィックな考え方が注目されている。また、ニューロンのグループがベイジアンネットワークを構成していることなどニューロンの大局的構造に注目した研究や、リカレントニューラルネットワーク（RNN）、LSTM（Long Short-Term Memory）、Reservoirコンピューティング[44]、オートエンコーダのような生物脳由来ではない工学的な情報表現法、学習法の研究の成果が上がりつつある。

米国エネルギー省（DoE）は、特に生物脳に忠実な再現を目指すのではなく、工学的に効果的であればよいと割り切って研究を進めているが、脳科学者の中には、現在のニューラルネットワークは、1950年代までの脳研究の成果に基づいているだけなので、もっと最近の成果を取り入れるべきだとする研究者も多い。新しい神経・脳モデルができれば、それに基づいて、様々な脳機能が再現される可能性があるが、同時に、忘れやすく、エラーを起こしやすく、疲れやすく、錯覚に陥りやすい脳が再現されてしまう可能性もある。

1.7.6.8 AIコンピュータの今後の研究開発の方向性

AIプロセッサは、ニューラルネットワークの単純なアーキテクチャを活かすことで、従来のロジック

※44
力学系的な定式化に基づくニューラルネットワークの一種。Echo State NetworkやLiquid-state Machineなどの種類がある。

CPUに比べて集積度を上げやすいこと、ReRAMやメモリスタのようなアナログ素子を使用することにより一つの素子が数ビットの情報量を蓄えることが可能となることから、従来のCPUにおける微細化の壁に打ち勝つことができる可能性を持っている。更に、ノイマン型アーキテクチャから非ノイマン型への変更により、メモリとプロセッサを小さく切り分けて近接した場所に置き、極端な高並列型のアーキテクチャとすることが可能である。これらの考えに基づき、大規模なニューラルネットワークを省電力で実装する試みが続けられている。

ディープラーニングにおける今後の課題の一つであるパターン認識と記号的処理（論理、言語、記号、計画、創造など）の解決に向けた研究開発においても、AIプロセッサの発展が寄与する可能性がある。現在のAIプロセッサは、認識以外の論理、言語、記号、計画、創造などを効率よく実行することができない。その解決の方向性は、脳科学・神経科学の研究からもたらされる示唆を工学的にモデル化した、脳とは異なる構造として実現する可能性もある。海馬、偏桃体、小脳など、大脳皮質とは異なる脳領域を合わせた超構造が必要になるかもしれない。脳の記憶は、海馬から入力されることは分かっているが、その記憶が大脳皮質のほかの領域に運ばれて固定化するメカニズムも解明されていない。

また、今までに作られてきたニューラルネットワークプロセッサは、順方向計算のみで、学習機能を持たない。現在のところ、学習に関してはサーバ側に任せるとする考えが強い。しかし、学習と認識処理を同時進行させられることに生物脳の特徴が隠されている可能性もある。このように、現在のニューラルネットワークを単純に大規模化することのみで良いのかどうかについては、不明なことが多い状況である。

ニューラルネットワークの応用領域は、急速に拡大しつつある。生物脳が学習によって様々な能力を発揮できるのと同様、AIプロセッサは、特に応用を限った発展を考える必要はない。しかし、AIプロセッサの開発を急ぐのであれば、まさにAIプロセッサの設計のための研究を高優先度のテーマとすべきであろう。そのような設計技術が生まれれば、あとは、次々と高性能のAIプロセッサが設計できるようになる。

1.7.6.9 量子計算機

更に長期的には、量子計算の動向にも注目しておく必要がある。量子計算の原理には主に三つに分類できる（**表11**）。

■表11　量子計算原理[45]

	量子コンピュータ	量子アニーラ	量子ニューラルネットワーク	
計算原理	状態ベクトルの ユニタリ回転（閉鎖系）	ハミルトニアンの断熱変化 （閉鎖系）	測定フィードバック系の 量子相転移（開放系）	
開発機関 ビット数 結線数	IBM ／ Google 5 ～ 9 ビット ≦ 10	D-WAVE 1,152 ビット 3,300	NII-Stanford 100 ビット 1 万	NTT 2,048 ビット 400 万
動作温度 物理系	極低温（$10m^{K}$） 超伝導量子回路	極低温（$10m^{K}$） 超伝導量子回路	室温（300^{K}） 光パラメトリック発振器 ネットワーク	
適用先	問題に隠れた 周期性がある場合 （暗号解読など）	NP 困難・NP 完全問題 （組合せ最適化など）	NP 困難・NP 完全問題 （組合せ最適化など）	
有効ビットの割合 解ける問題サイズ	— —	95% N ≦ 15～17	100% N ≦ 2,048	

2011年5月にD-Wave Systems（カナダ）が、世界初の商用量子コンピュータ「D-Wave One」を

※45
「光を使って難問を解く新しい量子計算原理を実現～量子ニューラルネットワークの開発～」科学技術振興機構ウェブサイト
<http://www.jst.go.jp/pr/announce/20161021/>

※46
“トポロジカル量子コンピュータ,” 日経サイエンスウェブサイト
<http://www.nikkei-science.com/page/magazine/0607/topology.html>

発表した。

D-Wave Oneは128量子ビットチップセット上で最適化問題を解くために、量子アニーリングを用いた量子コンピュータである。ただし、D-Waveの実装方法は、極低温を必要とする超伝導状態を用いており、現実性に乏しい。

Microsoftは2016年から、IBMやGoogle、D-waveとは全く違った戦略で量子コンピュータの開発に取り組んでいる。Microsoftは、世界中の共同研究者と一緒に量子コンピュータを構築するための取組を積極的に行っており、研究コンソーシアム的な研究施設Station Qを世界各国に建設し、量子コンピューティング[46]に関する研究を行っている。

対して、我が国の国立情報研究所とNTTで開発している量子ニューラルネットワークは、量子状態の重ね合わせのデバイスとして光を用いており、室温（300K=約27℃）での動作が可能な点で注目されている。現在約2,000ビットでの動作が報告されているが、今後の研究の更なる進展が期待される[47]。量子ニューラルネットワークの開発は、内閣府革新的研究開発推進プログラム（ImPACT）「量子人工脳を量子ネットワークでつなぐ高度知識社会基盤の実現」にて、2014年から2018年の計画で進められている。

海外の動向としては、2016年7月に量子情報科学と高性能コンピューティングの推進についての計画を発表[48]したほか、欧州では欧州委員会から2016年5月にQuantum Technologies Flagshipの立ち上げが発表され、10億ユーロ以上の大規模プロジェクトが立ち上がった[49]。

参考文献

[1] Shmuel Winograd, “Arithmetic complexity of computations,” Siam, vol.33.

[2] Olga Russakovsky et al., “ImageNet Large Scale Visual Recognition Challenge,” IJCV.

[3] Krizhevsky. A. et al., “ImageNet Classification with Deep Convolutional Neural Networks,” *NIPS 2012: Neural Information Processing Systems, Lake Tahoe, Nevada.*

[4] Maxime Oquab et al., “Learning and transferring mid-level image representions using convolutional neural networks,” CVPR2014.

[5] David Silver et al., “Mastering the game of Go with deep neural networks and tree search”, Nature 529, pp.484-489.

[6] Jeffrey Dean et al., “Large Scale Distributed Deep Networks,” *Advances in Neural Information Processing Systems 25（NIPS 2012）.*

[7] R. Zhao et al., “Accelerating Binarized Convolutional Neural Networks with Software Programmable FPGAs,” ISFPGA2017.

[8] Ben Varkey Benjamin et al., “Neurogrid: A Mixed-Analog-Digital Multichip System for Large-Scale Neural Simulations,” Proceedings of the IEEE, Vol.102, No.5, pp.699-716.

※47
「光を使って難問を解く新しい量子計算原理を実現～量子ニューラルネットワークの開発～」科学技術振興機構ウェブサイト
<http://www.jst.go.jp/pr/announce/20161021/>

※48
“Realizing the Potential of Quantum Information Science and Advancing High-Performance Computing.” White House Website
<https://obamawhitehouse.archives.gov/blog/2016/07/26/realizing-potential-quantum-information-science-and-advancing-high-performance>

※49
“European Commission will launch €1 billion quantum technologies flagship.” European Commission Website
<https://ec.europa.eu/digital-single-market/en/news/european-commission-will-launch-eu1-billion-quantum-technologies-flagship>

1.8 グランドチャレンジによる研究開発の推進

1.8.1 総論

現在までの人工知能（AI）の歴史の中で、チェス、将棋、囲碁などのゲーム、チューリングテスト（1.4.3項参照）、ロボットや自動運転など、様々なテーマで「グランドチャレンジ」が長期的な目標として設定されることにより、研究開発を推進する力を産み出してきた。グランドチャレンジは、大きな目標を掲げ、それを達成することで研究開発を加速するプロジェクト推進手法である。米国のアポロ計画に見られるように、「人類を月面に送り込み、安全に帰還させる」という明確かつチャレンジングな目標が設定される。同時に、その達成の過程で、社会的にも産業的にも重要な一連の技術が生み出されることがグランドチャレンジの設定において非常に重要である。

AIの分野では、人間に勝つことのできるチェスコンピュータの開発というチャレンジが、非常に早い段階から掲げられた。このチャレンジは、1997年にIBMの「DeepBlue」が、当時の世界チャンピンであるガルリ・カスパロフ（Garry Kasparov）氏に勝利することで達成されたが、その過程において、多くの探索アルゴリズムや並列計算技術など、広く普及している技術が生み出された。

その後、ゲーム題材として、将棋、囲碁、クイズショー、ポーカーなどがチャレンジの課題として設定され、実際に達成されてきた。例えば、将棋に関しては、情報処理学会の「コンピュータ将棋プロジェクト」[1]によって2015年にトッププロ棋士に追いつき、2017年には第二期電王戦が開催され、佐藤天彦名人と将棋ソフト「Ponanza」の対戦が行われてPonanzaが二局とも勝利した。

更に、我が国の研究者が中心となって1990年代半ばに始まったロボカップは「2050年までに、完全自律型のヒューマノイドロボットで、FIFAワールドカップの優勝チームとFIFAの公式ルールで試合を行い、勝利する」という目標を掲げ、世界45か国で数千人の研究者を巻き込む巨大プロジェクトとなっている。さらにロボカップへの参加者がKIVA Systems（現在のAmazon Robotics）、Aldebaran Robotics（現在のSoftbank Robotics）等の名だたるロボット企業を創業している[1]。

グランドチャレンジは、目標の達成自体が大きな社会的・産業的意義を持っている場合と、目標自体は、極めて難度が高く、インパクトがあるものの、それ自体には、直接的な社会的・産業的重要性は、必ずしも大きくはない場合とがある。後者の場合は、「ランドマーク型グランドチャレンジ」である。ロボカップを例に取ると、サッカーで世界チャンピオンになったとして、それ自体が、直接社会や産業の役に立つわけではない。しかし、その過程で生み出される技術が世の中に大きなインパクトを与えるというものである。

つまり、人類の歴史に残るような目標を掲げるが、それ自体は「記念碑（ランドマーク）」にすぎない。真の目標は、そこに到達する過程にあるということである。これは同時に、何をランドマークとして設定するかが最も重要であることを意味する。成功するランドマークプロジェクトは、次の三つの要件を満たすものである。

（1）社会・産業的に重要になりそうな一群の次世代技術の開発を要求する課題であること。
（2）その成功や進歩が、一般の人々にも分かる明確な形で示されること。
（3）最終目標が、歴史的記念碑となることが明白で、だからこそ困難が予想されるが、すぐに第一歩を踏み出すことは可能であること。

※1
「コンピュータ将棋プロジェクトの終了宣言」情報処理学会ウェブページ <http://www.ipsj.or.jp/50anv/shogi/20151011.html>

AIにおけるグランドチャレンジの多くは、ランドマーク型であり、そこからおびただしい技術が出され、世の中に普及している。

グランドチャレンジの成功と、最近のAIの発展を受けて、新たなグランドチャレンジを設定する動きが出てきている。システム・バイオロジー研究機構の北野宏明氏は、2050年までにノーベル賞級の科学的発見を可能とするAIを生命科学分野等で開発するというグランドチャレンジを提唱している[2]。また、公立はこだて未来大学の松原仁氏らは、ショートショートを創作させることを目標としたグランドチャレンジを提唱し、既に開始している。今後、これらの新しいグランドチャレンジが充実していくことにより、AIが加速していくことが期待される。

1.8.2 ゲームとAIの進化

ゲームは、ルールが明確である、勝ち負けによって手法の良し悪しが明確に評価できる、目標にできる強い人間が存在する、などAIの研究の題材として優れている。それにより多くのゲームを対象としたグランドチャレンジが行われてきた。

1.8.2.1 チェス

チェスは西欧で知性のシンボルとされているので、AIの例題としてのチェス（の世界チャンピオンに勝つコンピュータを開発すること）は、AIの研究が始まって約50年間ずっと中心的な例題となっていた。AIの最初のグランドチャレンジである。ジョン・マッカーシー（John McCarthy）氏は、チェスのことをAIの「ハエ」と称した。遺伝学が「ハエ」を題材として大きな進歩をしたように、AIはチェスを題材として大きな進歩をしたという意味である。チェスの「場合の数」はほぼ10^{120}である。ある局面でルール上指せる合法手の数を分岐数というが、チェスの平均分岐数は約35である。チェスは平均80手で勝負がつくので、35の80乗すなわち10^{120}が場合の数となる。

クロード・シャノン氏とアラン・チューリング氏は、チェスの探索にゲーム理論でジョン・フォン・ノイマンらが開発したミニマックス法[2]を使うことを提案し、このミニマックス法がその後のゲームの探索の基本となった。ゲームのプログラムを強くするには、

1）ミニマックス法を基本とした探索手法の改良
2）局面を点数化する（静的）評価関数の精緻化

の2つが求められる。理論的には例外が存在するものの、経験的にほとんどの場合、ゲームはより深く先読みしたほうが強くなるので、同じ時間でできるだけ深く先読みできる探索手法が望ましい。

チェスのプログラムを強くすることを目指して様々な工夫が試みられた。ミニマックス法は探索の末端の局面の全ての評価値をしらみつぶしで求めなくてはならないので、時間がかかってその分深く読めないという欠点がある。そこで、ミニマックス法と探索結果は同じでそれより効率がいい手法が経験的に開発された。それがアルファベータ法である。

チェスのプログラムで経験的に使われているのを、アルゴリズムとしてまとめたのがドナルド・クヌース（Donald Knuth）氏である。クヌース氏は末端の局面の数がN個のとき、アルファベータ法は最も効果が高い場合に$\sqrt{N}$個だけ評価値を求めればいいことを明らかにした。

※2
何らかの評価関数に基づき、最大の損失が最小になるように行動の意思決定を行う戦略。

アルファベータ法の効果が高くなるのは、展開した探索木が評価関数の値の大きい順番になっているときである。したがって、1手先を読むたびに評価値を計算して、大きい順に並べ替えておくのがいいことになる。チェスのプログラムで経験的にそのことが分かり、それがのちに反復深化（iterative deepening）という探索手法として定式化された。新しい探索手法の多くは、チェスを例題として開発されたと言ってよい。

例えば、ある指し手の評価値だけが他の評価値とかけ離れているときに、その指し手に注目して、その指し手だけをより深く読むという選択的深化（selective deepening）、評価値がどの程度信頼できるかを表わす共謀数（conspiracy number）とそれを拡張した証明数（proof number）・反証数（disproof number）などが有名である。

探索を効率的に行うためのハッシュ表[3]、ビットマップなど、データ構造の工夫もチェスを通して確立した。チェスは（アルファベータ法を使って）ルール上指せる全ての手を読むという全数探索が有効だったので、スーパーコンピュータやチェス専用マシンを使うことによって、探索の速度を上げようという試みが盛んになされた。並列に探索するアルゴリズムも、チェスを例題にして盛んに研究された（アルファベータ法は探索全体をアルファ値、ベータ値によって制御するので並列に探索するには困難があった）。

また、チェスは駒の再利用ルールがないので、ゲームの進行に伴って駒の数が単調に減少していく。駒が盤面に数個しか残っていない局面になると、コンピュータは（ほぼ）しらみつぶしの探索によってその局面を解く（双方が最善手を続けたら先手が勝つのか、後手が勝つのか引き分けになるかを求める）ことができる。この探索をあらかじめ行ってデータベース化したものが終盤データベースである。コンピュータはこれを持っていれば、このデータベースに含まれる局面で最善手を指すことができる。1980年代には盤面残り5駒の全ての局面の終盤データベースが作られた。1990年代から2000年代にかけて盤面残り6駒のほとんどの局面の終盤データベースが作られた（その間にコンピュータが世界チャンピオンに勝ってしまったので、終盤データベースを作る意味が薄くなったと言える）。その後も7駒の終盤データベースが作られている。

チェスは何度も人間との対戦を経たのちに1997年に「Deep Blue」が世界チャンピオンのカスパロフ氏に勝利した。6回戦で5回戦が終わった時点では1勝1敗3引き分けのイーブンであったが、最終戦でカスパロフ氏が緊張のあまり序盤で大悪手を指して負けてしまった。これはフロック勝ちで、この時点ではまだカスパロフ氏の実力はDeep Blueに勝っていたと思われる。

これでグランドチャレンジは目標を達成したことになる。その後の進歩によりコンピュータは人間より明らかに強くなっている。Deep Blueはスーパーコンピュータにチェス専用マシンを数百台並べた構成であったが、もはやパソコン 1 台でも人間が敵わないまでになっている。

1.8.2.2　将棋

チェスよりも場合の数が大きいゲームに中国将棋（10^{150}）、将棋（10^{220}）、囲碁（10^{360}）が存在する。中国将棋は探索問題として見るとチェスに近い（既に人間よりもコンピュータのほうが強くなっている）が、将棋と囲碁はチェスよりはるかに場合の数が大きく、チェスとは異なる手法が必要なので、チェスに続く例題として適切である。将棋はチェスと同じ敵の重要な駒（キングあるいは玉）を捕まえるゲームであるが、チェスは敵から取った駒が使えないのに対して、将棋では敵から取った駒

が再利用できる（「持ち駒」制度と呼ばれる）ため、終盤は序盤より分岐数が大きくなる。チェスは

※3
キーと値をペアで管理するデータ構造。

収束型ゲームであるが、将棋は発散型ゲームなのである。

将棋はチェスよりも場合の数がはるかに大きく、チェスで有効であった探索手法がそのままでは使えないので、チェスの次の探索研究のよい対象になった。将棋は日本固有のゲームなので、将棋を対象とした研究は当然のこととして日本が中心になった。このことが日本におけるゲーム情報学研究を活発にして、世界の中でゲーム情報学において日本が主要な立場を占める原動力になったと思われる。

将棋のプログラムの開発は1970年代に始まったが、当時のコンピュータの能力では将棋はチェスのような全数探索は無理だったので、前向き枝刈り[4]の探索手法が盛んに研究された。ミニマックス法（アルファベータ法）、反復深化などチェスで有効だった手法で将棋でも使える手法はおよそ全て使った。将棋は発散型ゲームなので、チェスで有効だった終盤データベースの手法は使えない。その代わりに詰め将棋という将棋から派生したパズルを解くアルゴリズムの研究が盛んになされた。

詰め将棋の研究は1990年前後から本格的に進められ、その中で様々な探索の手法が試された。有効だったのはチェスで提案された（そしてチェスではあまり有効でないとされた）証明数・反証数を用いた手法である。詰め将棋のプログラムは2000年前後には既にプロ棋士を超える能力を示した。

評価関数はチェス同様に手作業で作成と改良を行っていたが、チェスの評価関数は駒の損得という明快な基準があったものの、将棋の評価関数は複雑でなかなか強くならなかった。2000年代の半ばに登場した保木邦仁氏（現在、電気通信大学）の「Bonanza」（ボナンザ）によって、コンピュータ将棋は革命的な進歩を果たした。保木氏の工夫は、

（1）それまで将棋は前向き枝刈りの探索をしていたのをチェスのように全数探索にした。
（2）それまで評価関数は手作業で作っていたのを棋譜からの機械学習で作るようにした。

の2点である。この工夫をしたボナンザが圧倒的な強さでコンピュータ将棋のトップに立ったので、ほかの研究者・開発者もこぞってこれらの方法を取り入れた。

特に上記の（2）の方法は強豪のプログラム全てが取り入れており、「ボナンザメソッド」と呼ばれている。保木氏はBonanzaのアルゴリズムをすぐに公開し、またプログラムのソースコードも無償で公開した。これはコンピュータチェスの文化を引き継いだものであるが、研究成果を公開するという習慣がこの研究領域の発展を支えているものと思われる。

2010年代になって、コンピュータとプロ棋士が対戦するようになった。2013年、2014年と電王戦と称してプロ棋士5人とプログラム5つが対戦したが、3勝1敗1分け、4勝1敗と、ともにコンピュータが圧勝した。この時点で既にコンピュータはトップクラスのプロ棋士（竜王、名人）のレベルに達した。

現在のトップ棋士のシンボルである羽生善治氏との対戦はすぐに実現しないと思われるので、情報処理学会は2015年10月に将棋で人間とコンピュータの強さを問うことは学問的には結論が出たという終了宣言を行った。事実上グランドチャレンジは2015年をもって目標が達成されたことになる。その後2017年にPonanzaが佐藤天彦名人に圧勝している。

1.8.2.3　囲碁

囲碁は中国発祥のゲームであるが、中国では廃れて日本で盛んになった（いま中国で盛んになったのはいわば日本からの逆輸入である）。囲碁は、ほかに似たルールのゲームが存在しない、漢字を使っていないので親しみやすい、などの理由で世界的に普及している。

※4
見込みのなさそうな手を試行しないことで、手を読む数を減らす方法。

最初にコンピュータ囲碁の研究がなされたのは1960年代である（チェスよりは遅いが将棋より早い）。囲碁もチェスのように探索によって次の手を決めようとしたが、囲碁の場合の数は10^{360}とチェス（や将棋）よりはるかに大きく、普通の探索によっていい手を見つけるには候補手が多すぎて強くならなかった。2000年代になっても、まだとても弱い状態であった（初心者レベルよりはましでもせいぜい初級者レベルであった）。

囲碁も将棋のボナンザメソッドのような革命的な手法が現れた。それがモンテカルロ木探索である。この元となったモンテカルロ法はフォン・ノイマン氏の命名といわれるシミュレーションによって解を求める方法である。1990年代にこれを囲碁に適用するというアイディアが発表されたものの、そのときは成功しなかった。2000年代になってレミ・クーロン（Remi Coulom）氏が「Crazy Stone」という囲碁プログラムのなかでモンテカルロ法を応用したモンテカルロ木探索を採用し、このCrazy Stoneが圧倒的な強さを示した。

囲碁にモンテカルロ法を適用するということは、ある局面から白と黒が交互にランダムに終局まで打ち進めるというシミュレーションを多数行って、勝つ確率が一番高い手を選ぶということである。そこには囲碁の知識はほとんど何もはいっていない。この一見単純な方法で強くなることに驚き、その後の囲碁プログラムはみんなこの方法を取り入れている。それで囲碁プログラムは一気にアマチュアの六段程度の実力に達した[5]。

最近までは日本の「ZEN」（これもモンテカルロ木探索を用いている）が、Crazy Stoneを抜いて最も強い囲碁プログラムであった。これらのプログラムはまだ互先（ハンディなし）で戦うのは無理であるが、トッププロ棋士と4子（初期局面に4個の石をあらかじめ置く）のハンディで勝つまでになっていた。トッププロ棋士に勝つのはまだ10年はかかると思われた。

そこに2016年1月、Googleの「AlphaGo」というプログラムが二段のプロ棋士に互先で5戦5勝の成績を挙げたと発表して大ニュースになった。Alpha–Goは、

1）ディープラーニング
2）モンテカルロ木探索
3）強化学習

という三つの手法をうまく組み合わせている。大量のプロ棋士の棋譜をデータとしてディープラーニングによってある程度の強さのプログラムを作り、そのプログラム同士の強化学習によって更に強くした。これまでコンピュータ囲碁で成功しなかった評価関数を、実質的に作ったことがAlphaGoの大きな特徴である。手を決める部分では、従来手法であるモンテカルロ木探索を使っている。

その後、2016年3月に、AlphaGoは韓国のトッププロ棋士のイ・セドル（Lee Sedol）氏と対戦して4勝1敗で圧勝した。AlphaGoの改良版である「Master」は2016年末から2017年初めにかけて（持ち時間が短い早碁ではあるが）、世界中のトッププロ棋士相手に60勝で負けなしという成績を収めた。中国の「Fine Art」やZENの改良版である「DeepZenGo」も、ディープラーニングを取り入れてトッププロ棋士といい勝負をするまでになった。囲碁も一気に2016年から2017年に、グランドチャレンジの目標が達成されたことになる。

※5
美添一樹ほか『コンピュータ囲碁-モンテカルロ法の理論と実践－』
共立出版.

1.8.3 ロボカップ

ロボカップは（RoboCup）、「2050年までに、完全自律型のヒューマノイドロボットで、FIFAワールドカップの優勝チームとFIFAの公式ルールで試合を行い、勝利する」という目標を掲げているロボットとAIのグランドチャレンジの一つである。ロボカップは、浅田稔氏、野田五十樹氏、北野宏明氏、松原仁氏ら、日本人研究者が中心となって1990年代初めに構想され、1997年に名古屋で第１回大会が開かれた。運営母体となるThe RoboCup Federationは、スイスに登録されている非営利組織である。

最初は、サッカーに関するリーグから始まったが、すぐに災害救助（ロボカップレスキュー）や教育（ロボカップジュニア）に関する活動が加えられた。現在では、家庭用ロボット（ロボカップ@ホーム）や物流など産業用途ロボット（ロボカップインダストリアル）をタスクとしたリーグが増えてきている（**表12**）。これらのタスクは、理事会で承認され、技術的マイルストーンの検討、研究上と産業上の有用性などから審査され、適切とみなされると追加される。現在、45か国から、数千人の研究者が参加し、教育では数十万人の子供たちが参加する、ロボットとAI分野における世界最大のプロジェクトである。

■表12　ロボカップのリーグ構成（2017）[6]

ロボカップのリーグ種別	概要
ロボカップサッカー	サッカー
ヒューマノイドリーグ	人型ロボット
中型ロボットリーグ	車輪中型
小型ロボットリーグ	車輪小型
シミュレーションリーグ	シミュレーション
ロボカップレスキュー	災害救助
レスキュー実機リーグ	災害救助ロボット
レスキューシミュレーションリーグ	シミュレーション
ロボカップジュニア	教育
ロボカップ@ホーム	家庭内向けロボット
ロボカップインダストリアル	産業用途ロボット
ロジスティクスリーグ	物流ロボット
@ワーク	オフィス環境向け
アマゾン・ロボティクス・チャレンジ	倉庫内物流ロボット

また「リーグ」と呼ばれる各々のカテゴリは、毎年技術要件や競技規則が見直され、最終的にFIFAの正式ルールと一致するようにマイルストーン管理がなされている。例えば、2014年には、**図45**の様に、フィールドの広さ、周囲の設定なども含め飛躍的に難易度が高い技術要件をクリアする必要がある競技規則となっている。

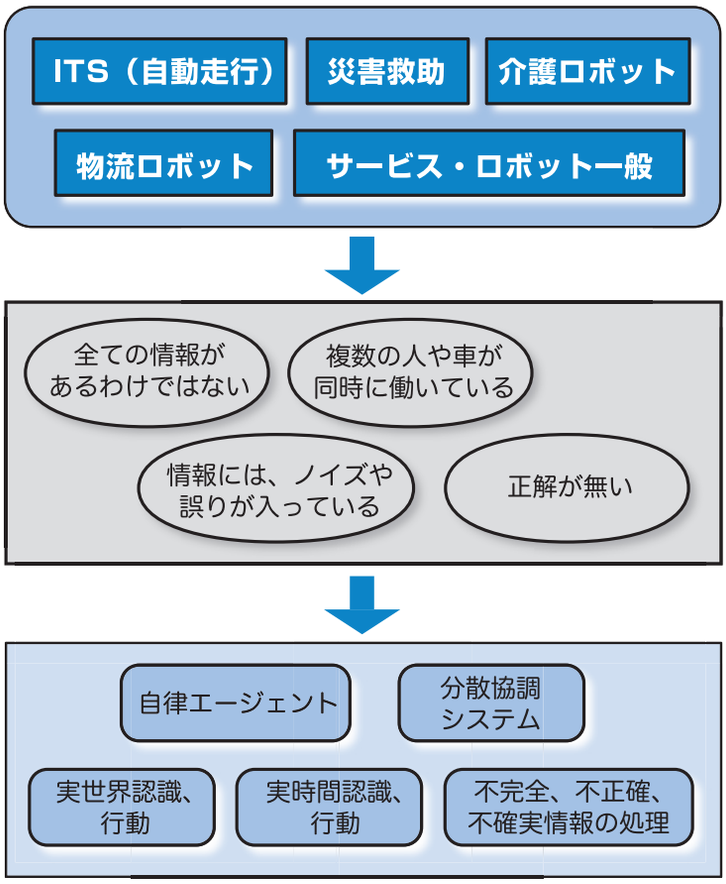
■図45　中型ロボットリーグ

ロボカップがサッカーを題材とした理由は、ロボット工学とAIの分野で21世紀中頃に、重要な応用領域（自動走行、物流ロボットなど）を想定して、それらの応用領域の特徴（不完全

※6
「RoboCup 2017 Nagoya Japan」ロボカップ2017ウェブサイト
<http://www.robocup2017.org/index.html>

情報[7]など）を抽出した。その上で、一連の基幹技術になりそうな項目を同定し（自律エージェント、分散協調システム、実時間システム、不完全情報下での意思決定システムなど）、それらを包含し、誰にでも一言で理解してもらうことが可能で、更に研究者自身が熱くなれるテーマとして選ばれた（**図46**）。

これから必要となる技術は、チェスや将棋のように、全ての状況が理解でき、順番に駒を動かすような問題ではない。不確実な情報を基に刻一刻と変化する状況下で、ベストではないかもしれないがベターな判断を下し、それを実行できる技術体系だろう、という分析であった。いろいろなテーマの候補があったが、最終的に次世代技術の要素を最も含んでいて、世界中で受入れられるテーマとして、サッカーが選ばれている。



■図46　ロボカップのタスク設定のロジック

ロボカップを通じて開発された技術を基盤に、起業し、それが大きな成功を収める事例も出てきている。ロボカップの小型リーグを通じて開発された技術を基礎に設立された会社（KIVA Systems、米国）が、2012年に、7億7千万ドル（約800億円）という大きな評価額でAmazonに買収された。更に、ロボカップの標準プラットフォームリーグにワンメイクのヒューマノイドロボットを提供していた、フランスのAldebaran Roboticsが、ソフトバンクに1億ドルの出資を受けるということが起きたのである。

KIVA Systemsは、コーネル大学（米国）のチームを率いたラファエロ・ダンドレア（Raffaello D'Andrea、現在チューリッヒ工科大学）氏らが、ロボカップ向けに開発した技術をベースに、パッケージングから倉庫内の物品移動、発送までも自動化するロボットシステムを開発して事業化した会社である。ちなみに、コーネル大学チームは、ロボカップの小型ロボットリーグに1999年から2003年まで参加し、4回の優勝を飾っている。ダンドレア氏は、そのときのコーネル大学チーム（Cornell Big Reds）のリーダーである。

サッカーロボットでは、複雑でしかも状況が変化する環境下で自律的に目的の場所に移動する、障害物との衝突を回避する、味方のロボットと連携する、といった機能が必須である。これらの機能を大規模オンラインショップ向けに設計し、トータルソリューションを実現したのである。

ロボカップレスキューは、ロボカップの目標がサッカーであり、しかもその達成時期を2050年とかなり先に設定しているため、より早い段階で世の中に還元できる取組も必要であるとの認識から始まっている。

※7
意思決定時において、必要な情報が不完全であること。

災害救助という目的に最適化するため、サッカーロボットと違い、完全自律である必要はない。実際には、操縦者が遠隔操作できる上に、ある程度の自律制御で探索効率を上げる方式が実際的である。ただし、直接ロボットの見えるところから操作することはできない。なぜなら、災害現場では、ロボットからかなり離れた場所から操縦することが想定されるからである。

この背景には、阪神・淡路大震災でロボット工学が無力だったという反省があり、日本のロボット関係者の間の議論から、計画は急速に動き出した。2001年8月にシアトルで第1回の大会が開催された。

その1か月後、9.11のテロが発生した。米国から参加していた南フロリダ大学のチームは、大会後、レスキューロボットを遠征用にパッキングしたまま休暇に入り、新学期を迎えるところだったが、テロ発生の知らせに、このロボットを車に積んでニューヨークの現場に入った。ロボカップレスキューは、その構想段階から米国の連邦緊急事態管理局（FEMA）などとも交流を深めていたこともあり、即時に現場の救助活動に統合され、2週間にわたり探索活動の一翼を担い、その有効性は高く評価された。

このような実績が、1999年からロボカップに関心を寄せていた米国防衛先端研究計画局（DARPA）の強い興味を引きつけ、ロボカップのノウハウを利用しながらのDARPA Grand Challenge設立へと結び付いている。

この段階での研究は、阪神・淡路大震災、オクラホマの連邦ビル爆破、9.11同時多発テロ、トルコでの一連の地震などによる被害での救援活動を想定していたため、ビルの倒壊現場などで、瓦礫の間隙から中に入り、被災者を発見するシナリオで開発されている。同時に、新潟県中越地震などで問題となった土砂崩れなどには、無力であることも認識されていた。

東日本大震災においても、津波が被害をもたらした大きな原因であり、レスキューロボットの有効性は限定的になった。もっとも、福島第一原発に投入されている国産ロボット（千葉工業大学未来ロボット技術研究センターが中心に開発）は、ロボカップレスキューの2007年大会に運動性能の部で優勝したシステムをベースに開発されている。ロボカップレスキューは、救助ロボット開発の手法としての有効性は確認されたが、広範かつ複雑な災害現場へのレスキューロボットの有効的な投入には、更に現実的な設定に近づけると同時に、ロボットが最も有効な局面への集中的な課題設定も必要となる。

ロボカップには、このほかにも、各々の目標を設定したリーグが存在し、本来のグランドチャレンジの手法を更に広範に援用した、コンペティション駆動型研究開発のプラットフォームへと変貌を遂げて進化している。

1.8.4 DARPAにおけるグランドチャレンジ

DARPA（アメリカ国防高等研究計画局）とは、米国防総省の研究開発機関であり、インターネットの起源となるARPANETやGPSを開発したことで有名である。DARPAはこれまでに様々なグランドチャレンジと呼べる研究開発プロジェクトを実施しており、なかでも有名なのは「DARPA Grand Challenge」（2004年、2005年）や「DARPA Urban Challenge」（2007年）、「DRC」（DARPA Robotics Challenge）である。

DARPA Grand Challenge、DARPA Urban Challengeはいずれも自動運転技術を競い合う競技会で、DARPA Grand Challengeでは未舗装路を走破する技術、DARPA Urban Challengeでは市街地の交通ルールを守りながら走破する技術が競い合われた。本項では2012年から2015年にかけて行われたDRCを紹介する。

DRCは2011年に発生した東日本大震災をきっかけにして、プログラムマネージャーであるギル・プラット（Gill Pratt）氏が立案・実施した競技会形式の研究開発プログラムであり、その目的は災害発生時に人を支援できるロボットシステムを開発することである。参加者は以下の四つのトラックを選択し

て参加することが可能であった。

トラックA：DARPAからの予算支援を受けて、ハードウェア及びソフトウェアの全てを開発する
トラックB：DARPAから予算支援を受けてソフトウェアを開発し、後述のVRCで勝ち残ればハードウェアプラットフォームの提供を受けて開発を継続する
トラックC：DARPAから予算支援は受けずにソフトウェアを開発し、VRCで勝ち残ればハードウェアプラットフォームの提供を受けて開発を継続する
トラックD：DARPAからの支援は受けずにハードウェア及びソフトウェアの全てを開発する

トラックDのような参加形態が設定されたのは、DARPAが軍関連の組織であり、そこからの支援を受けることに対して抵抗のある組織が多いことが要因であると考えられる。

競技会は以下に示す3度にわたって行われ、ハードウェアプラットフォームとして米Boston Dynamicsが開発したヒューマノイドロボットであるAtlas、シミュレーションプラットフォームとして米OSRF（Open Source Robotics Foundation）が開発したGazeboが提供された。

（1）ハードウェアプラットフォームの提供を受ける参加者を決定するための、コンピュータシミュレーションによる競技会Virtual Robotics Challenge（VRC、2013年6月）
（2）決勝戦へ進むチームを決定するためのDRC Trials（2013年12月）
（3）決勝戦であるDRC Finals（2015年6月）

DRC Trialsでは東京大学出身の若手研究者が立ち上げたベンチャーSCHAFTがトラックAで参加し、優勝を収めた。同社はTrialsの直前に上記Boston Dynamics等とともにGoogleによって買収されていたことと合わせて、関係者の間では非常に大きな話題となった。

DRC Finalsの競技の概要は次のとおりであった。

- 8つのタスクを連続して実行し、完了できたタスクの数が多いもの、タスクの数が同じ場合はより短い時間で完了できたものが高成績となる。8つの競技とは（1）車両を運転する、（2）車両から降りる、（3）ドアをあけて室内に入る、（4）バルブを回す、（5）工具を持ち、壁に穴をあける、（6）サプライズタスク、（7）不整地を移動する、又は障害路を通過する、（8）階段を登る、である。
- 競技時間は1時間
- ロボットは無線で動作しなければならない（外部電源なし、転倒防止索なし）

DRCではロボットの自律性を高める研究開発を促進するため、通信制限がルールに盛り込まれた。ロボットとオペレータの間の通信路は2種類あり、一つは通信速度が9,600bpsと非常に遅いが、常につながっており、双方向通信が可能な通信路。もう一つは通信速度が300Mbpsと速いが、屋内エリアに入ると通信が途切れ途切れとなり、最大で30秒の通信遮断が発生し、更に情報はロボットからオペレータへの一方向でしか送れない通信路である。オペレータが画像を見ながらレスポンスのよい遠隔操作を行うためには、高いバンド幅の通信路が必要であり、これを制限することによってロボットの自律性を高める研究が行われるように誘導している。

DRC Finalsには全世界から23チームが参加し、日本からは5チームがトラックDで参加した。新エネルギー・産業技術総合開発機構（NEDO）の支援を受けて参加したAIST-NEDO（産業技術総合研究所）、NEDO-JSK（東京大学稲葉研究室）、NEDO-Hydra（東京大学中村研究室、千葉工業大学、大阪大学、

神戸大学）の3チームとHRP2-Tokyo（東京大学稲葉研究室）、Aeroである。Finalsの結果は次表（**表13**）のとおりとなった。

優勝したのは韓国のTeam KAISTであり、使用機体であるDRC-Huboは人型でありながら膝と爪先部分に車輪を持ち、平坦なところでは正座のような姿勢で高速かつ安定に移動し、階段の移動や作業時に立ち上がって作業を行った。日本チームは10位のAIST-NEDOが最高位となった。

■表13　DRC Finalsの競技結果

国	チーム	ポイント	時間	移動機構	ロボットのタイプ
韓	TEAM KAIST	8	44:28:00	2脚 / 車輪	DRC-HUBO
米	TEAM IHMC ROBOTICS	8	50:26:00	2脚	ATLAS
米	TARTAN REASCUE	8	55:15:00	4脚 / クローラ	独自
独	TEAM NIMBRO RESCUE	7	34:00:00	4脚 / 車輪	独自
米	TEAM ROBOSIMIAN	7	47:59:00	4脚 / 車輪	独自
米	TEAM MIT	7	50:25:00	2脚	ATLAS
米	TEAM WPI-CMU	7	56:06:00	2脚	ATLAS
米	TEAM DRC-HUBO AT UNLV	6	57:41:00	2脚 / 車輪	DRC-HUBO
米	TEAM TRAC LABS	5	49:00:00	2脚	ATLAS
日	TEAM AIST-NEDO	5	52:30:00	2脚	HRP-2
日	TEAM NEDO-JSK	4	58:39:00	2脚	独自
韓	TEAM SNU	4	59:33:00	2脚	ROBOTIS
米	TEAM THOR	3	27:47:00	2脚	ROBOTIS
日	TEAM HRP2-TOKYO	3	30:06:00	2脚	HRP-2
韓	TEAM ROBOTIS	3	30:23:00	2脚	ROBOTIS
米	TEAM VIGIR	3	48:49:00	2脚	ATLAS
伊	TEAM WALK-MAN	2	36:35:00	2脚	独自
米	TEAM TROOPER	2	42:32:00	2脚	ATLAS
独	TEAM HECTOR	1	2:44	2脚	ROBOTIS
米	TEAM VALOR	0	0:00	2脚	独自
日	TEAM AERO	0	0:00	4脚	独自
米	TEAM GRIT	0	0:00	4脚	独自
香港	TEAM HKU	0	0:00	2脚	ATLAS

DRCは人型ロボットのみを対象とした競技会ではなかったが、Atlasが人型であったこと、階段等脚でなければ移動が困難な環境が含まれていたことから、多くのチームが人型のロボットで競技に臨んだ。しかし結果を見ると、2脚以外の移動機構を採用したチームが上位に集中している。2脚の移動機構を持つロボットのほぼ全てが一度は競技中に転倒したことも合わせて考えると、二足歩行は更なる技術開発が必要である。

DRCは災害時に人に代わって活躍できるロボットを開発することを目的として実施されたが、優勝したTeam KAISTですら8つのタスクを実施するのに45分を要した。仮に同じタスクを人が実施していれば5分程で完了するものと思われ、迅速な対応が求められる災害現場にロボットを投入するには不十分である。自動運転の技術はDARPA Urban Challengeから10年を経て実用化に漕ぎ着けており、災害対応ロボットに関しても実用化に向けて研究開発を継続していくことが重要である。

1.8.5 AIによる科学的発見に関するグランドチャレンジ

AIによる科学的発見は、一つの大きな分野である。この分野でのグランドチャレンジとして、ロボカップの提唱者の一人でもあるソニーコンピュータサイエンス研究所（Sony CSL）の北野宏明氏は、「2050年までにノーベル賞級の科学的発見を行うAIシステムを開発する」という目標を掲げたグランドチャレンジを提唱している。特に、医学生理学賞をターゲットとしている。

更に、ノーベル賞は人間に与えられる賞であることから、「Nobel Turing Challenge」として、ノーベル賞級の科学的発見をするAIシステムが、選考委員会からAIであると見破られないで受賞をするというチャレンジを課している。

今までも、AIシステムによる科学法則の発見に関する研究は行われてきた。しかしながら、それらの研究は、既に発見されている法則を、計算機で再発見できるかという試みや、エキスパートシステムの一種であるなど、本当の意味で大きな科学的発見に結び付く展開にはならなかった。

しかし、現在多くの科学分野で大規模データを扱うことが一般化し、膨大な計算を可能とする各種のインフラストラクチャが実現している。同時に、1990年代中頃から登場したシステムバイオロジーの分野では、大規模網羅データを系統的に測定する技術を加速すると同時に、詳細な生命の設計原理や分子機構への洞察を深めた。この状況の変化は、新たにAIによる科学的発見という分野に、再度、グランドチャレンジを設定して、取り組むべき時期にきたと思われる。

この一つの作業仮説は、科学的発見とは、大規模仮説空間の生成・探索と、それらの仮説の高速検証であるというものである。この作業仮説の背景には、今までのグランドチャレンジは、大規模データ、大規模計算、更には機械学習という三つの要因で成り立っていたという分析がある。であるならば、科学的発見も、大規模仮説空間の生成と探索で可能であろうと思える。

このチャレンジを実現するためには、一連のプラットフォームの構築が必要である。このため、まず、各種のデータ並びにモデル表現などに関して標準化を行うコミュニティを成立させている。さらに、解析ソフトウェアなどの相互運用性を実現する必要がある。そこで、「Garuda Platform[8]」を構築し、これらの問題を解決しようとしている。これらの基盤があって初めて、極めて大きな科学的発見を行うAIシステムの開発が可能であると考えている。また、生物実験の精度を向上させ、効率を追求したロボット実験システムの開発も行われている。これらの流れが連動し、このグランドチャレンジを成功に導くと思われる。

このグランドチャレンジは、グローバルな分散協調プロジェクトとなると思われる。仮想的な大規模プロジェクトをどう進行させるのかという新たなマネージメント上のチャレンジでもある。しかし、各々のチャレンジは、極めて重要かつ新規性の大きなものであり、グランドチャレンジ達成への中間段階で、大きな成果の展開も期待できる。

参考文献
[1] 「ノーベル賞級の発見をするAI　人の限界を超えた科学研究へ」日経エレクトロニクス, 2016.7, pp.97-108.
[2] Hiroaki Kitano, “Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine for Scientific Discovery,” *AI Magazine*, vol.37 No.1, pp.39-49.

※8
Sony CSL北野宏明が開発したシステムバイオロジー研究のための
統合型データ解析プラットフォーム。

1.9 各国の研究開発の現状

1.9.1 総論

我が国では、経済産業省、総務省、文部科学省にそれぞれ人工知能（AI）研究のためのセンター（産業技術総合研究所、情報通信研究機構（NICT）、理化学研究所）があり、それぞれAI研究を推進するとともに、連携して研究開発に当たることとなっている。民間企業においても、自動運転や生産ロボットなどの一部の業界において、本格的な研究開発に取り組み始めている状況である。我が国のAIに関わる研究開発の今後の発展に向けて、アルゴリズムの基礎研究、応用研究を更に振興するとともに、ロボティクスや計算用のデバイスなどものづくりの強みを活かした研究開発が有効と考えられる。

海外については、主に米国の情報系企業のディープラーニングに関する取組が早い段階から展開されている。2006年にディープラーニングの研究の発端となった論文を執筆したジェフリー・ヒントン氏は、トロント大学（カナダ）の教授であるが、現在Google（米国）と兼任している。そのほか、ディープラーニング分野の著名な研究者の多くは、Microsoft Research（米国）や、Facebook AI Research（米国）等のいち早く設立された民間情報系企業の研究所に移籍や兼任、アドバイザリー等の形で関わっており、情報系企業のAIの研究開発戦略を担っている。また、中国はアカデミックの研究のほか、情報系企業もAIに力を入れており、追い上げが著しい状況である。

米国政府は、「Preparing for the future of artificial intelligence」というレポートを2016年5月に発表し、AIの研究開発の方向性を示した。もともと米国では2015年にイノベーション戦略“A STRATEGY FOR AMERICAN INNOVATION”を策定し、ニューロサイエンス、コネクテッドカーや自動運転車、先進マニファクチャリング、スマートシティといったAIに関連の深いテーマを重点分野として指定しているほか、BRAIN Initiative、Precision Medicine Initiativeや国防高等研究計画局（DARPA）でのSyNAPSEプログラム等、AIと関連の深い分野の研究開発も推進しており、今後も様々な分野でAIへの投資を継続すると予想される。

英国はケンブリッジ大学、オックスフォード大学において従来からAI研究が盛んであるとともに、先端的なディープラーニングの研究開発を行っているGoogle DeepMindの本拠地でもある。ドイツも、ドイツ人工知能研究センター（DFKI）において民間企業との共同研究を多く実施しており、マックスプランク研究所や大学等を含めて、AIや脳科学まで含めれば研究人材が一定数存在している。ディープラーニングへの対応は米国に先を越されたと言わざるを得ないが、従来から「Horizon 2020」[1]において、ヒューマン・ブレイン・プロジェクトを実施しており、脳のシミュレーションから応用まで幅広い範囲で研究を推進してきている実績もある。今後のAI研究に脳科学の知見が取り込まれる過程で重要な寄与をする可能性がある。

中国では、人材の豊富さも手伝って、AIの研究開発が加速している。2016年3月には政府は「インターネットプラス　AI3年行動実施法案」を発表し、自動運転、ロボット、スマートホーム等の重点分野において世界トップクラスの企業を育成する目標を掲げている。また、Baidu、Alibaba、Tencent等の情報系企業もAIへの対応を急いでいる。

※1
2014年～2020年まで7年間にわたって、EUの研究開発を促進するためのプログラム。

1.9.2 各国の政策・プロジェクトの現状

1.9.2.1 我が国のAI研究開発政策

我が国では、「第5期科学技術基本計画」※（平成28年1月22日閣議決定）において、AIを「超スマート社会」を実現するための競争力向上のための基盤技術として位置付け、その強化を推進することとなった。超スマート社会とは、「必要なもの・サービスを、必要な人に、必要な時に、必要なだけ提供し、社会の様々なニーズにきめ細かに対応でき、あらゆる人が質の高いサービスを受けられ、年齢、性別、地域、言語といった様々な違いを乗り越え、活き活きと快適に暮らすことのできる社会」である。ICTを最大限に活用し、サイバー空間とフィジカル空間（現実世界）とを融合させた取組により、人々に豊かさをもたらす「超スマート社会」を未来社会の姿として共有。その実現に向けた一連の取組を更に深化させつつ「Society 5.0」として強力に推進し、世界に先駆けて超スマート社会を実現していくこととされた。

第5回「未来投資に向けた官民対話」（平成28年4月12日）において、安倍総理がAIの研究開発目標と産業化のロードマップを平成28年度中に策定することを表明した。それを受けて、AIの研究開発・イノベーション政策の司令塔となる「人工知能技術戦略会議」が平成28年4月18日に発足し、総務省、文部科学省、経済産業省の3省が連携してAI技術の研究開発と成果の社会実装の加速に当たることとなった。

人工知能技術戦略会議の下には、上記3省のそれぞれが所管するAI研究のためのセンターが存在するため、各センターの研究の総合調整を行う場として研究連携会議が設置されるとともに、人材育成、標準化・ロードマップ作成、技術・知財動向分析、規制改革等のテーマについて研究開発と産業の連携総合調整を図る産業連携会議が設置されて議論が行われている[2]。平成29年3月31日には「人工知能技術戦略」[3]を公表するとともに、「人工知能の研究開発目標と産業化のロードマップ」[4]が策定された。この中では、「生産性」、「健康、医療・介護」、「空間の移動」の3分野及び横断的分野として「情報セキュリティ」が重点分野とされ、3センターが連携して研究開発に取り組むとともに、産学官が有するデータ及びツール群の環境整備を実施することとされた（**表14**）。

■表14　3センターの連携による研究開発テーマ

重点分野	研究テーマ概要
生産性	ハイパーカスタマイゼーションの実現を目指し、消費者の需要を反映させた適時適量・多品種少量生産を可能とする次世代生産技術の研究開発。
健康、医療・介護	予防医療の高度化による病気にならないヘルスケアの実現を目指し、認知症を含む疾患の早期発見、最適な治療法選択、対処を可能とするシステムの研究開発。
空間の移動	SIP における自動走行システムと連携しながら、地図データの意味付けやユニバーサルコミュニケーション技術による移動空間の高付加価値化を実現するスマートモビリティの研究開発。

※2
人工知能技術戦略会議「資料1　人工知能技術戦略会議について」新エネルギー・産業技術総合開発機構ウェブサイト
<http://www.nedo.go.jp/content/100790387.pdf>
※3
人工知能技術戦略会議「人工知能技術戦略（人工知能技術戦略会議とりまとめ）」新エネルギー・産業技術総合開発機構ウェブサイト
<http://www.nedo.go.jp/content/100862413.pdf>
※4
人工知能技術戦略会議「人工知能の研究開発目標と 産業化のロードマップ」新エネルギー・産業技術総合開発機構ウェブサイト
<http://www.nedo.go.jp/content/100862412.pdf>
※5
産業構造審議会　産業技術環境分科会　研究開発・イノベーション小委員会「イノベーションを推進するための取組について」 経済産業省ウェブサイト <http://www.meti.go.jp/committee/sankoushin/sangyougijutsu/kenkyu_kaihatsu_innovation/pdf/report01_01.pdf>

経済産業省の産業構造審議会 産業技術環境分科会 研究開発・イノベーション小委員会ではイノベーションを推進するための取組について議論が行われた。平成28年5月13日に公表した中間とりまとめ[5]では、AIの産業構造を一変させ得る技術として位置付け、国費による国家プロジェクトの研究成果の一部であるデータについて、オープンイノベーションによる利活用を促進するためのデータ戦略を検討することも重要とされた。

新エネルギー・産業技術総合開発機構（NEDO）では、平成27年度から「次世代人工知能・ロボット中核技術開発」をスタートし、次世代AI技術分野として①計算論的神経科学の知見を取り入れた脳型AIやデータ駆動型のAIと知識駆動型のAIの融合を目指した研究開発、②様々な次世代AI技術を統合するフレームワークの研究開発、③標準的ベンチマークやデータセットの整備、革新的ロボット要素技術分野として④革新的なセンシング技術、⑤革新的なアクチュエータ技術、及び⑥ロボットインテグレーション技術の研究開発を実施している。また、平成29年度からは、⑦社会実装に関する研究開発の先導研究を実施する予定としている（**表15、表16**）。また、「IoT推進のための横断技術開発プロジェクト」では、アナログ型抵抗変化素子を用いた脳型推論集積システムの開発や、革新的アニーリングマシンの研究開発等を実施している[6]。

■表15　次世代人工知能・ロボット中核技術開発」の次世代人工知能技術分野の研究開発項目

No.	研究開発項目	概要
①	大規模目的基礎研究・先端技術研究開発	最新の計算論的神経科学の知見を取り入れた脳型 AI 及びデータ駆動型の AI と知識駆動型の AI の融合を目指すデータ・知識融合型 AI に関して、大規模なデータを用いた実世界の課題への適用とその結果の評価を前提とした目的基礎研究（大規模目的基礎研究）と、世界トップレベルの性能の達成を目指す先端技術の研究開発を実施する。
②	次世代人工知能フレームワーク研究・先進中核モジュール研究開発	広範な AI 応用の研究開発や社会的実用化に資するため、研究開発項目①の成果である脳型 AI 技術、データ・知識融合型 AI 技術、そのほか大学や企業が保有する様々な AI 技術をモジュール化し統合するための次世代 AI フレームワークと、次世代 AI 技術を統合し、多様な応用に迅速につなげるための核となる先進中核モジュールの研究開発を実施する。
③	次世代人工知能共通基盤技術研究開発	次世代 AI の共通基盤技術として、AI 技術の有効性や信頼性を定量的に評価し、性能を保証するための方法、そのために必要となる標準的問題設定や標準的ベンチマークデータセット等が満たすべき性質と構築の方法に関する研究開発を実施する。また、それらを用いて、研究開発項目①、②の成果の評価を行う。
⑦	次世代人工知能技術の社会実装に関するグローバル研究開発（平成 29 年度より実施）	次世代 AI 技術の社会実装が求められる領域として、「人工知能の研究開発目標と産業化のロードマップ」における当面の検討課題のうち、（1）生産性、（2）健康、医療・介護、（3）空間の移動の 3 領域において、関連する課題の解決に資するため、次世代 AI 技術の社会実装に関する研究開発を先導研究から実施する。なお、AI 技術とものづくり技術との融合等を国内外の叡智を結集して、グローバルに行うことを考慮する。

■表16　次世代人工知能・ロボット中核技術開発」の革新的ロボット要素技術分野の研究開発項目

No.	研究開発項目	概要
①	革新的なセンシング技術（スーパーセンシング）	屋外等の外乱の多い空間でも、的確に信号抽出ができる画期的な視覚・聴覚・力触覚・嗅覚・加速度センシングシステムやセンサと行動を連携させて、検知能力を向上させる行動センシング技術等の研究開発を実施する。
②	革新的なアクチュエーション技術（スマートアクチュエーション）	人共存型ロボットに活用可能なソフトアクチュエータ（人工筋肉）、高度な位置制御やトルク制御を組み合わせてソフトウェア的に関節の柔軟性を実現する新方式の制御技術や機構等の研究開発を実施する。
③	革新的なロボットインテグレーション技術	実環境の変化を瞬時に認知判断し、即座に対応して適応的に行動する技術や個別に開発された要素技術を効果的に連携させ統合動作させるシステム統合化技術等の研究開発を実施する。

※6
「IoT推進のための横断技術開発プロジェクト」新エネルギー・産業技術総合開発機構ウェブサイト
<http://www.nedo.go.jp/activities/ZZJP_100123.html>

■表17　AI研究センター(AIRC)のチーム構成

研究チーム	研究の概要
知識情報研究チーム	データに内包される意味を理解し、知識を抽出する技術の研究を実施。文章形式のデータのみならず様々な形式のデータを分析し、その中に記述されている出来事の因果関係や、登場する言葉の概念構造、情報の鮮度と客観性、情報間の矛盾などを、AIが認識し、データベース化する技術。
確率モデリングチーム	様々なデバイスから得られる実世界の大規模データ(ビッグデータ)と、人が持つ知識の両方を融合し、高度なタスクを実行するAIを学習させる確率モデリング技術の開発。
脳型人工知能研究チーム	大脳皮質に関する神経科学的知見をヒントにしたBESOMと呼ぶ機械学習アルゴリズム（ベイジアンネット、自己組織化マップ、独立成分分析などを組み合わせたもの）の実用化を目指した研究。
機械学習研究チーム	ベイジアンモデリング、カーネル法、ディープラーニングなどの先端的な機械学習技術の理論基盤、アルゴリズムの研究開発から、リモートセンシングデータ、医療データ、経済データ、ロボットの感覚・運動データ等の実データへの応用まで幅広く研究を実施。
人工知能クラウド研究チーム	実世界から取得される多種多様大量のデータ(ビッグデータ)を対象とした高度かつ高性能なデータ処理技術の確立と、これを基盤として、AI技術の容易かつ迅速な適用を可能にする次世代AIフレームワークの実現を目指した研究。
人工知能応用研究チーム	機械学習に基づく画像解析や音響データ解析による異常検知などをコア技術とし、社会インフラ診断及び医療診断・ヘルスケア支援に資する技術の実用化に向けた研究。
サービスインテリジェンス研究チーム	人々が主体的・共創的にインテリジェンス（観察、判断、行動力）を高める方法論とそれを効率的に実現するAI技術を研究。具体的には介護、看護、健康増進、保育、教育、理美容などの現場に知識工学、設計工学、データ工学、認知科学、バイオメカニクスなどを適用し、横展開可能な技術を開発する。
計算社会知能研究チーム	人と人、人とサービスの相互作用を取り入れたシステム設計を工学的に支援するため、人々の振る舞いを継続的にセンシングする技術と、人を系に組み込んだシミュレーション手法を組み合わせ、サービス導入、改変の影響を都市規模で予見することを目指した研究。
地理情報科学研究チーム	多種多様かつ膨大な地理空間情報を知的に処理できる基盤を開発し、科学研究だけでなく環境管理、資源開発、防災といった具体的な応用に結び付けた研究。
生活知能研究チーム	多様な生活機能変化者に適合した安全な生活、自立した生活、高度な社会参加のある生活の実現といった社会的インパクトのある具体的課題を設定し、IoT技術、画像処理技術、生活データベース技術、ロボット技術などの研究を推進。また、大規模生活データからニューノーマル化した生活課題をいち早く見つけ、そのソリューションを開発可能にする「生活知識循環エコシステム」の創造も長期的な狙いとしている。
オーミクス情報研究チーム	ライフサイエンスにて生産される大量のオーミクス情報から有用な知識を抽出するための情報解析技術及びAI技術の開発。具体的には、ゲノム及びエピゲノム情報解析技術、遺伝子ネットワーク解析技術、タンパク質立体構造シミュレーション技術などの開発を実施。
インテリジェントバイオインフォマティクスチーム	ゲノム情報を始めとする多様で膨大な生命情報に関するデータから生体分子に関する知識発見を行うためのバイオインフォマティクス技術の開発と、疾病因子の推定や生体分子の機能解析などを通じた創薬などへの応用。

経済産業省が所管する産業技術総合研究所では、平成27年5月1日にAI研究センター（AIRC）を設立した（**表17**）。主要な目的基礎研究として、①人間の脳の情報処理原理に関する最新の神経科学の知見を包括的に取り入れた人間の脳に近い脳型AIと、②実世界の大量のデータに基づくデータ駆動型のAIとウェブ上の大規模な知識グラフなどに基づく論理的・形式的な知識駆動型のAIの2つを融合して、大量かつ多様な実世界のデータを深く理解し、人間の意思決定を支援するデータ・知識融合型AIの研究を行うことを目標としている（**図47**）。

これらの目標のため、AIフレームワーク上で要素技術を統合した先進中核モジュールを実装して、製造業やサービス産業などの幅広い分野での産学連携による実サービスから得られる大規模なデータを使った実証研究、研究用データセットなど、AI技術の研究の基盤となるリソースの整備の実施を通じて、幅広い用途でのAI技術の有用性を提示し、産業競争力の強化と豊かな社会の実現に貢献することを目指している。

具体的には、「人工知能に関するグローバル研究拠点整備事業」[7]では医療・介護現場、住環境、工場等の模擬環境の整備と個別分野のデータの収集・管理、解析、2次提供を行うデータ基盤の構築等を実施するオープンイノベーション・ハブ拠点を構築している。また、「人工知能・IoTの研究開発加速のための環境整備事業」[8]や、「人工知能処理向け大規模・省電力クラウド基盤」（ABCI）などの整備を進めており、ディープラーニングの研究開発の基盤となることが期待されている（1.7.3項参照）。

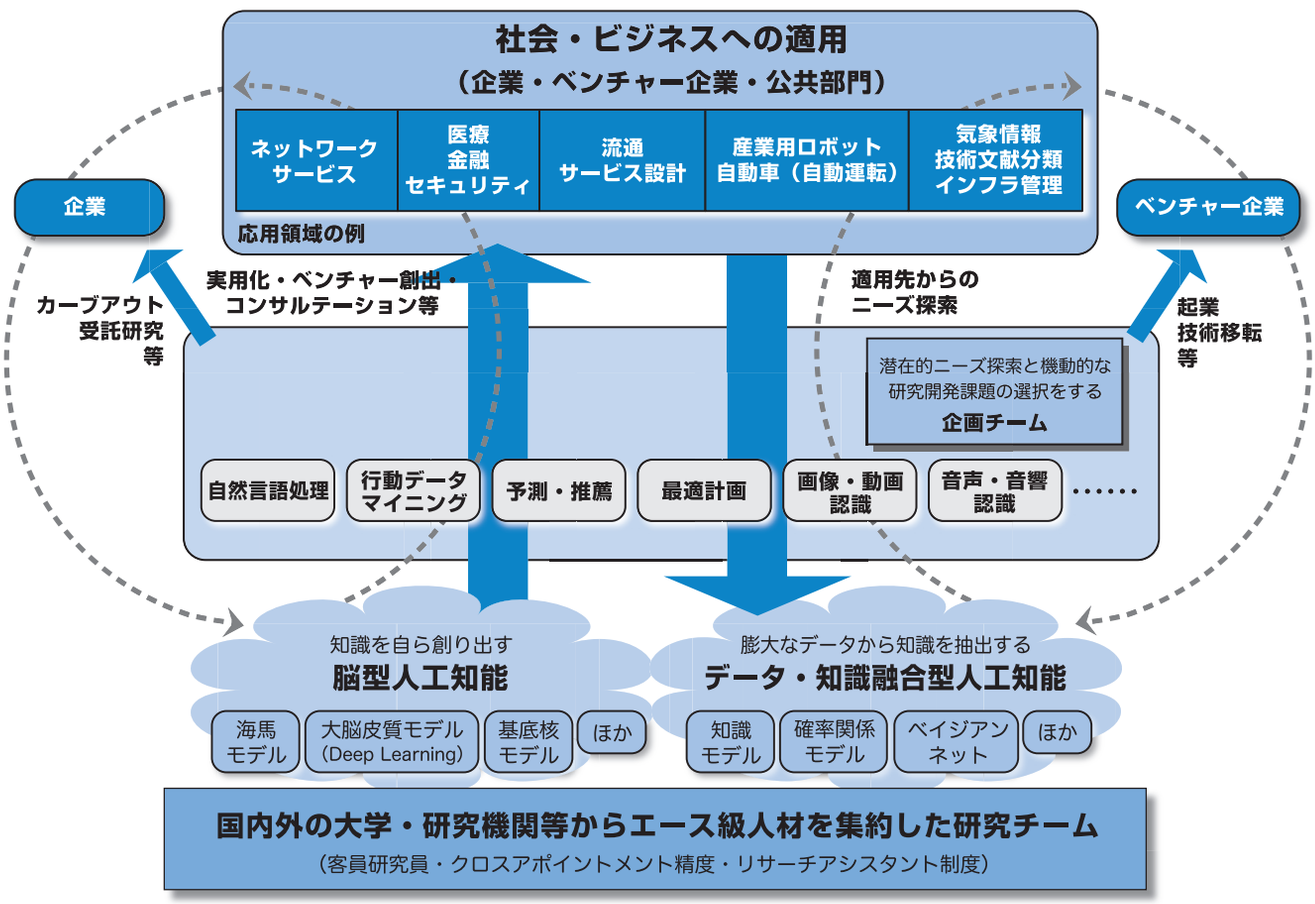
総務省では、総務大臣の諮問機関である「情報通信審議会情報通信技術分科会技術戦略委員会」において、平成28年7月に「次世代人工知能推進戦略」を取りまとめた[10]。本戦略では、我が国で注力していくべき研究開発分野として、8個のテーマが掲げられている（**表18**）。

※7
産業構造審議会　産業技術環境分科会　研究開発・イノベーション小委員会「イノベーションを推進するための取組について」
経済産業省ウェブサイト <http://www.meti.go.jp/committee/sankoushin/sangyougijutsu/kenkyu_kaihatsu_innovation/pdf/report01_01.pdf>

※8
「IoT推進のための横断技術開発プロジェクト」新エネルギー・産業技術総合開発機構ウェブサイト
<http://www.nedo.go.jp/activities/ZZJP_100123.html>

■表18　次世代人工知能推進戦略」の研究開発テーマ

No.	概要
①	小規模データしか得られない場合に、強化学習やスパースモデリングと呼ばれる技法を用いて学習を実現する AI 技術の開発。
②	ディープラーニングの欠点（問題が複雑な場合に汎化能力が欠如するという本質的な課題や、入力と出力の関係がブラックボックスとなってしまい、システムに不具合が生じてもその原因の究明や品質保証が困難になる課題）を克服した機械学習法の研究開発。
③	少量のデータしか得られない場合でも、多数の入力データを活用することによって汎化能力が高められる半教師あり学習など新たな機械学習法の研究開発。
④	ロボット等の運動と AI の組合せにより、プランニングを行ったり、シンボルグラウンディングを行う問題に関する研究開発。
⑤	機械翻訳や音声翻訳などの自然言語処理技術と対訳コーパスの開発及び蓄積、並びにウェブや SNS、更には学術論文や公的文書等の多種多様な知識を利用する技術、こうした知識をより効率よく人間に伝え活用するための手段として対話ロボット等の開発。
⑥	ネットワーク上のクラウド等と自律的に処理を分担するとともに、システム間での情報共有が可能となる等、相互に通信し連携しながら自律的に判断、行動し、人の意思決定や行動を支援するための、IoT/ビッグデータ、AI を前提としたネットワーク型 AI 社会基盤の実現（例えば、異なる機械学習アルゴリズムの融合に基づいた通信の効率化や、情報のスパース符号化による通信量の削減、更には脳の動的なネットワークの再構成を模倣した効率よいルーティングなど）。
⑦	脳活動計測データ自体の解析への AI の適用。
⑧	人間の脳の情報処理メカニズムを参考にしたディープラーニングの新たなパラダイムの創出など、脳科学の知見の AI への適用。



■図47　AI研究センター（AIRC）における研究開発の取組[9]

※9
「「人工知能研究センター」を設立　－人工知能研究のプラットフォーム形成をめざして－」産業技術総合研究所ウェブサイト <http://www.aist.go.jp/aist_j/news/pr20150507.html >編集部作成

※10
「次世代人工知能推進戦略」　総務省ウェブサイト <http://www.soumu.go.jp/main_content/000424360.pdf >

※11
「ALAGIN 言語資源・音声資源サイト」ALAGIN 言語資源・音声資源サイトウェブサイト<https://alaginrc.nict.go.jp/>

※12
先進的音声翻訳研究開発推進センターウェブサイト<http://astrec.nict.go.jp/research/index.html>

※13
「「AIPプロジェクト（人工知能/ビッグデータ/IoT/サイバーセキュリティ統合プロジェクト）」に係る平成28年度戦略目標の決定について」文部科学省ウェブサイト <http://www.mext.go.jp/b_menu/houdou/28/05/1371147.htm>

所管のNICTでは、脳情報通信、音声認識、多言語音声翻訳、社会知解析、革新的ネットワーク技術等の研究開発をかねてより進めている。例えば、高度言語情報統合フォーラム（ALAGIN）[11]では、自然言語処理の研究に資する言語資源・音声資源の整備を実施している。また、脳情報通信融合研究センターでは、システム神経科学、情報通信技術、ブレインマシンインターフェース、ニューロイメージング技術及びロボット工学の研究を実施している。更に、先進的音声翻訳研究開発推進センターでは、東京オリンピック・パラリンピック競技大会が開催される2020年までに、国内の鉄道などの交通機関やショッピング施設、観光地、医療の現場などで活用される実用性の高い多言語音声翻訳技術や、企業などにおいて他国の特許を自動で翻訳できる多言語テキスト翻訳技術などを開発している[12]。

次世代人工知能推進戦略では、このようなNICTがこれまで整備を進めてきた言語情報データや脳情報モデルを基盤として、全国規模で利用可能とする「最先端AIデータテストベッド」の整備、脳機能に学び知能を理解・創造する次世代AI技術の研究開発、IoT／ビッグデータ／AI情報通信プラットフォームの開発等を推進することとしている。

文部科学省は「人工知能／ビッグデータ／IoT／サイバーセキュリティ統合プロジェクト」（AIPプロジェクト）を推進しており[13]、その研究開発拠点として、理化学研究所に革新知能統合研究センター（AIP）を平成28年4月に新たに設置した（**表19**）。

■**表19　AIPプロジェクトにおける研究テーマ**[14]

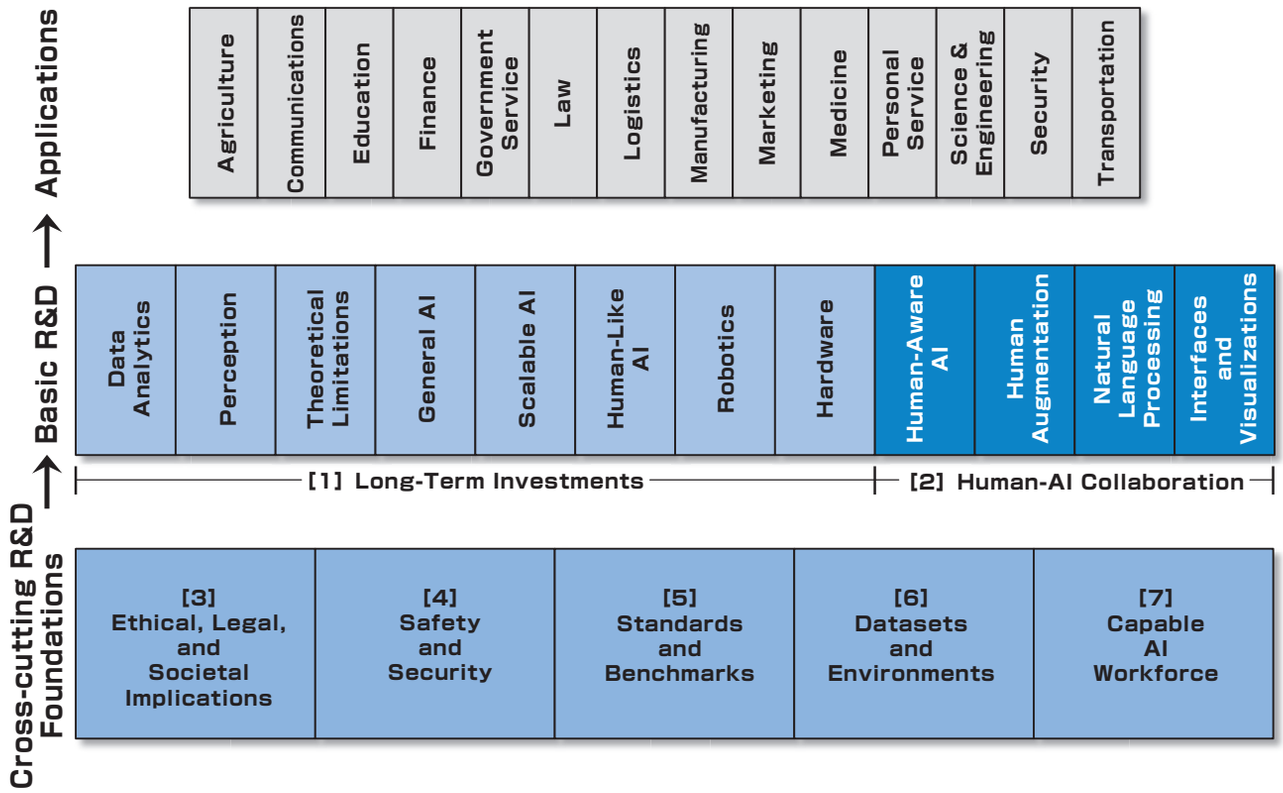
No.	概要
①	社会・経済等に貢献するため、多種・膨大な情報を組み合わせ解析する技術開発（カプセル内視鏡やCTなどから取得される膨大な医療画像を診断において高速処理する技術、電子カルテの高度解析による投薬や治療計画最適化をサポートする技術、病気の予兆を発見する技術等）。
②	多種・膨大な情報に基づき、状況に応じ最適化されるシステムのための技術開発（自動運転に関わる膨大な情報から安全走行に必要な情報のみを取捨選択し計算負荷を大幅に低減するデータ処理技術、災害発生時にネットワークを状況に応じ自律的に構成する技術、データの意味を高度に理解してデータの統合分析を可能とするオントロジー、時系列データのリアルタイム分析技術等）
③	多種多様な要素で構成される複雑なシステムに適用可能なセキュリティ技術開発（予測型セキュリティ技術、軽量暗号化アルゴリズムの開発・実装、セキュリティ・バイ・デザイン、来歴等のエビデンス情報（プロヴェナンス）によるデータ信頼性検証技術等）。

科学技術振興機構（JST）では、戦略的創造研究推進事業（新技術創出）の一環として、AIPに対応した研究領域を設定しプロジェクトを推進している。CRESTの新領域「イノベーション創発に資する人工知能基盤技術の創出と統合化」が平成28年度に新たに設定され、実社会の様々な分野に資するセンサ技術、実時間ビッグデータを扱うデータベース技術、システムセキュリティ技術、機械学習を核とするシステム最適化技術等の高度化と実世界データを総合的に実時間で処理し理解する統合化技術の研究開発が実施されている。

1.9.2.2　米国のAI研究開発政策

米国では、全米科学財団（NSF）、国防総省傘下のDARPAや、分野別には国立衛生研究所（NIH）、エネルギー省等により、AIや脳科学の技術開発投資が多く行われてきた。

※14
「急速に高度化・複雑化が進む人工知能基盤技術を用いて多種膨大な情報の利活用を可能とする統合化技術の創出」文部科学省ウェブサイト <http://www.mext.go.jp/b_menu/houdou/28/05/attach/1371148.htm>
※15
“A STRATEGY FOR AMERICAN INNOVATION.” The white house Website <https://obamawhitehouse.archives.gov/sites/default/files/strategy_for_american_innovation_october_2015.pdf>
※16
「ライフサイエンスのフロンティア―研究開発の動向と生命倫理―（平成27年度　科学技術に関する調査プロジェクト）」国立国会図書館ウェブサイト <http://www.ndl.go.jp/jp/diet/publication/document/2016/index.html>
※17
“Artificial Intelligence: CALO.” SRI international Website <https://www.sri.com/work/timeline-innovation/timeline.php?tag=security-and-defense#!&innovation=artificial-intelligence-calo>



■図48　米国のAIのR&D戦略の構成

2015年に国家経済会議（National Economic Council; NEC）と大統領科学技術政策局（OSTP）により策定された「A STRATEGY FOR AMERICAN INNOVATION」[15]では、ニューロサイエンス、コネクテッドカーや自動運転車、先進マニファクチャリング、スマートシティといったAIに関連の深いテーマを重点分野として指定している。

NSFは、2006年から多種・大容量のデータ処理等関連技術の基盤となる研究開発を継続的に支援しており、2015年からの新たなプログラムでは、基礎研究（3年）、学際研究（3～4年）、大規模研究（4～5年）の募集が数十万～100万ドル規模で実施されている。

2014年に発表されたBRAIN（Brain Research through Advancing Innovative Neurotechnologies）Initiative（BRAINイニシアティブ）では、脳内の神経回路構造を細胞から脳全体のレベルまで全ての階層について調べることで、脳の構造と機能、動作原理の解明を目指した研究が進められている[16]。

米国におけるAI研究では、歴史的にDARPAが果たしてきた役割が大きい。例えば、Appleの音声アシスタント「Siri」に利用されている技術の源流はDARPAの「CALO」（Cognitive Assistant that

※18
"Deputy Secretary of Defense Speech." U.S. Department of Defense Website <https://www.defense.gov/News/Speeches/Speech-View/Article/634214/cnas-defense-forum>

※19
"THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN,H October 2016. The white house Website <https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf>

※20
「ライフサイエンスのフロンティア―研究開発の動向と生命倫理―（平成27年度　科学技術に関する調査プロジェクト）」国立国会図書館ウェブサイト <http://www.ndl.go.jp/jp/diet/publication/document/2016/index.html>

Learns and Organizes)”プロジェクトにおいてSRI Internationalが開発した技術である[17]。近年では、自動運転を目指すアーバンチャレンジや災害用ロボットの実現を目指すロボティクスチャレンジ等のグランドチャレンジや（1.8節参照)、IBMのニューロモーフィックチップであるTrueNorthがDARPAのプロジェクトにより開発されている。

最近では、軍事装備品の諸外国に対する優位性を維持するための「第3のオフセット戦略」の中で、技術的に重要であるビルディングブロックとして、自律的なディープラーニングを用いたシステム、人と機械の協働、人の行動のエンハンスメント、人と機械の協調動作、半自律的な武器、の5点を挙げており、装備品へのAIの活用を目指すプロジェクトを活発化している[18]。

2016年12月にはオバマ政権が「The National Artificial Intelligence Research and Development strategic Plan」を発表しており[19]、長期的な基礎研究の継続性の重要性の指摘している。また、特に重要なテーマとして、人とAIのコラボレーションに資する技術を取り上げている（**図48**）。

1.9.2.3　欧州のAI研究開発政策

EUでは、2013年からの10年間にわたる重点科学プロジェクトとしてヒューマンブレインプロジェクト（HBP）が推進されており、総額12億ユーロが投じられる予定となっている[20]。プロジェクトのスコープは幅広く、脳に関係する遺伝子やたんぱく質の発現情報の取得から人の認知機能のメカニズムの解明、脳のシミュレーション、治療やロボットの研究のためのツール開発まで幅広い範囲の研究開発がなされている。

また、英国では、「UK Digital Strategy」が文化・メディア・スポーツ省により2017年3月に発表され、サイバーセキュリティ、フィンテック、ゲーム、仮想現実、GovTech[21]等におけるAIの活用を推進することにより、関連領域のビジネスを後押しするとしている[22]。

ドイツでは製造業の産業競争力強化を目指して「Industrie 4.0」が推進され、関連したシステム研究開発等を実施している。マックスプランク研究所や大学等を含めて、AIや脳科学まで含めれば研究人材が一定数存在している。ディープラーニングへの対応は米国に先を越されたと言わざるを得ないが、従来からHorizon 2020において、HBPに参画している研究者が多数おり、脳のシミュレーションから応用まで幅広い範囲で研究を推進してきている実績もあり、今後AIに脳科学の知見が取り込まれている過程で重要な寄与をする可能性がある。

また、Horizon 2020の下では、機械翻訳のプロジェクト「Quality Translation21」（QT21）等が実施されている。

1.9.2.4　中国のAI研究開発政策

中国では、人材の豊富さも手伝って、AIの研究開発が猛烈な勢いで加速している。2016年3月には政府は「インターネットプラス　AI3年行動実施法案」を発表し、自動運転、ロボット、スマートホーム等の重点分野において世界トップクラスの企業を育成し、2018年までに1000億元の市場を創出する目標を掲げている。

※21
政府、行政、公共分野を支援するテクノロジーのこと。
※22
政府、行政、公共分野を支援するテクノロジーのこと。 GOV. UK Website <https://www.gov.uk/government/publications/uk-digital-strategy>
※23
中華人民共和国中央人民政府ウェブサイト<http://www.gov.cn/xinwen/2017-02/20/content_5169236.htm>
※24
Yunhe Pan, “Heading toward Artificial Intelligence 2.0,” Engineering, vol.2 No.4, 2016, pp.409-413. Engineering Website <http://engineering.org.cn/EN/10.1016/J.ENG.2016.04.018 >

また、「科学技術イノベーション2030」の重点項目に最近「artificial intelligence 2.0（AI2.0）」が追加された[23]。AI2.0では、今後、ビッグデータに基づくAI、クラウドベースのAI、人と機械をハイブリッドすることにより実現するAI、自律的システムの4点を重視するとしている[24]。

1.9.3 民間企業の研究開発の現状

1.9.3.1 我が国の民間企業における研究開発動向

我が国の民間企業においては、特にディープラーニングに対する取組について、米国の情報系企業を中心とする取組に比べ、国内のAI関連の人材不足もあり（2.3節参照）、大手企業において国内外の研究機関と連携することで研究開発を進めている事例が見られる。例えば、Toyota Research Institute（米国）はシリコンバレーに研究所を開設し、スタンフォード大学等との共同研究を実施している。ソニーは、強化学習分野で世界的に著名な研究者であるマーク・リング（Mark Ring）氏がCEOを務めるスタートアップCogitai（米国）に資本参加し、ディープラーニングを用いた強化学習や予測・検知技術の応用に取り組んでいる。また、リクルートホールディングスは、データ分析の自動化技術を開発しているDataRobot（米国）に出資している。パナソニックは、大阪大学と人工知能技術とそのビジネス応用に関する人材開発を共同で行うための人工知能共同講座を開始している。

このほかに、AIの研究を実施している大手企業は、製造業や情報通信業の企業に多く見られる。東芝は、送電線の損傷等の異常画像をディープラーニングで生成し、学習用データの不足を補う仕組みの開発等を実施している。ホンダは、傘下のホンダ・リサーチ・インスティチュート・ジャパンにおいて、脳の計算原理のモデル化に取り組んでいるほか、「人と協調する人工知能技術」等を研究する研究開発組織を新設している。富士通研究所では、手書き文字認識について、教師データが少なくても学習可能なディープラーニング技術の開発を実施している。NECは、顔認証技術にディープラーニングを取り入れ、顔の向きの変化等にも強い、動画を対象とした顔認証を開発した。NTTは、コミュニケーション基礎科学研究所において、ディープラーニングを用いた音声認識技術の研究を実施している。

そして、日立製作所がイジング計算機と呼ばれる組合せ最適化問題に特化したチップを開発しているほか、ディープラーニングを高速かつ高効率に行うチップの開発が、パナソニック、デンソー、NEC、東芝、富士通等によって進められている（1.7.6項参照）。なお、ディープラーニングのオープンソースのフレームワークであるChainerを開発するプリファードネットワークスは、ファナック、DeNA、トヨタ自動車等と協業している（1.2.10項参照）。

1.9.3.2 米国の民間企業における研究開発動向

2006年にディープラーニングの興隆の発端となった論文を執筆したヒントン氏は、トロント大学の教授であるが、現在はGoogleと兼任している。このように、米国の場合、大学と民間企業の研究所の間の人材交流が大変活発であることが特徴である。ヒントン氏のほかにも、Facebook AI Researchのヤン・ルカン氏等、AI分野の著名な研究者が多く民間情報系企業の研究所に移籍や兼任、アドバイザリー等の形で関わっており、情報系企業のAIの研究開発戦略を担っている。

ジェフリー・ディーン（Jeffrey Dean）氏が率いるGoogle Researchは、1200名程度（2017年4月現在）の研究者を擁する。特にディープラーニングに特化したGoogle Brain Teamは、ディープラーニングの理論的研究から自然言語処理、機械翻訳等の研究を進めており、Neural Turing Machine等、パターンと記号処理の融合を目指す分野で本質的な研究を多く行っている。また、ディープラーニングのフレームワークであるTensorFlowをオープンソースとして公開している。

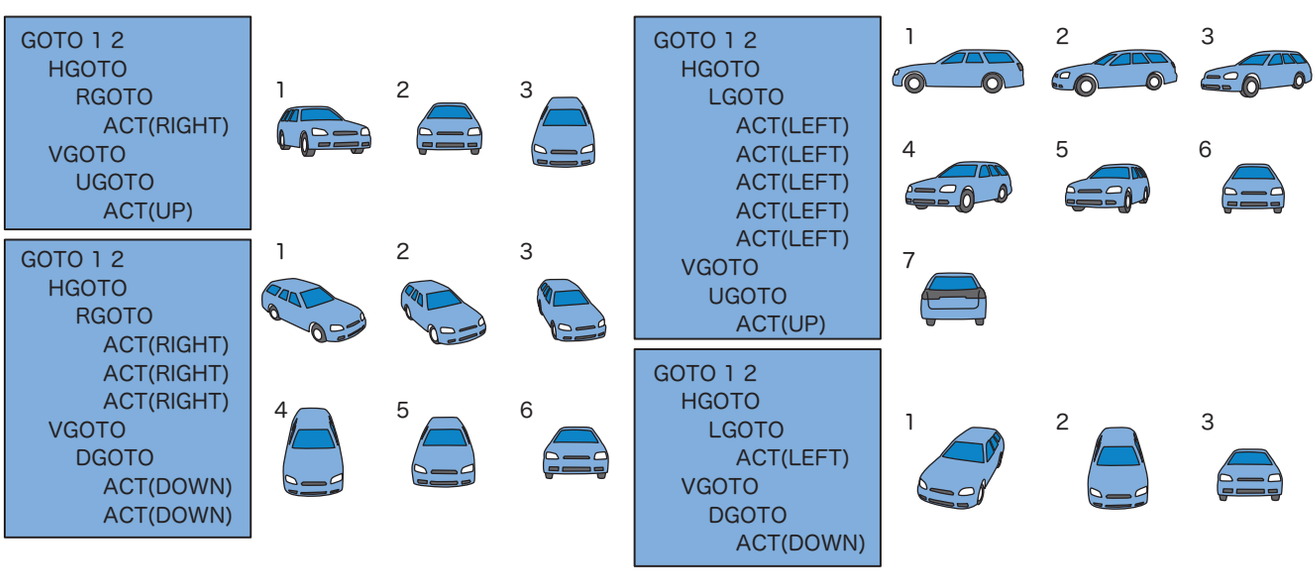
Vicarious（米国）は、AIのコア技術を開発するベンチャー企業であり、スタンフォード大学のAI Labのディレクターであるフェイ・フェイ・リー（Fei-Fei Li）氏や、スパースコーディングで有名なカリフォルニア大学バークレー校のブルーノ・オルシャウセン（Bruno Olshausen）氏が科学アドバイザーを務めている。注力する研究テーマとして、脳科学の知見を取り入れたニューラルネットワーク、少量のデータからの学習や教師なし学習など、今後のAIの研究開発でキーとなり得るものを挙げており、注目に値する（**図49**）。

少量のデータからの学習モデルの開発	・通常の深層学習では学習に大規模なデータセットが必要であるが、人間は1例で学習することも可能である ・出来るだけ少ないデータで学習可能なことが知能のコアである
教師なし学習	・深層学習の成功例は現在のところ教師あり学習 ・教師なし学習が多くの問題で重要となる
脳科学からの知見	・新皮質の構造によりアルゴリズムにもたらされる制約が脳の効率的処理の根本にある ・脳に関する知見を取り入れたアルゴリズムの開発を実施する
ネットワーク構造の重要性を強調	・現在の主流モデル（CNN）にとらわれず、compositionality を重視したネットワーク構造を採用
生成モデル	・「想像」することが出来るシステム ・因果的意味論を取り入れることが可能

■図49　Vicariousが掲げる研究テーマ

1.9.3.3.　欧州の民間企業における研究開発動向

英国はケンブリッジ大学、オックスフォード大学において従来からAI研究が盛んであるとともに、Google DeepMindの本拠地でもある。ディープラーニングと強化学習の組合



■図50　仮想3次元空間内に自動車の3Dモデルをランダムな向きで置き、その状態から正面を向く様に回すプログラムを生成する例[26]

※26
Scott Reed and Nando de Freitas, “Neural Programmer-Interpreters.” Cornell University Library Website <https://arxiv.org/abs/1511.06279>

※27
“Opening a new chapter of my work in AI.” Medium.com Website <https://medium.com/@andrewng/opening-a-new-chapter-of-my-work-in-ai-c6a4d1595d7b>

※28
“A Chinese Internet Giant Enters the AI Race.” MIT Technology Review Website <https://www.technologyreview.com/s/603070/a-chinese-internet-giant-enters-the-ai-race/>; Tencent AI Lab Website <http://ai.tencent.com/ailab/>

※29
検索日は2017年3月27日。

※30
欧州特許庁（European Patent Office, EPO）に出願され発行されたもの。

せで囲碁においてトッププロ棋士に勝利したAlphaGoを作ったGoogle DeepMindの研究開発動向は、仮想的な3次元の迷路を解くエージェントの開発や、リカレントニューラルネットワーク（RNN）でプログラミングを行ってしまう例など、今後のAI開発の方向性を見据えた取組を多く実施している（**図50**）。

ドイツも、DFKIにおいて、Volkswagen（ドイツ）など民間企業との共同研究を多く実施しており、自然言語処理、知識処理、仮想現実、データマイニング等の研究を実施している。

1.9.3.4　中国の民間企業における研究開発動向

中国では、Baidu、Alibaba、Tencent等の情報系企業のAIへの取組が先行している。Baiduでは、1300名程度がAI事業に関わっており、そのうち300名程度が研究者である[27]。

Tencentでは、50人の研究者を擁し、機械学習、コンピュータビジョン、音声認識、自然言語処理の研究を行っている[28]。ソーシャルアプリであるQQのデータを背景として、

高性能なビッグデータ解析のためのプラットフォームであるAngelを2017年の第一四半期にオープンソースとすることを発表している。

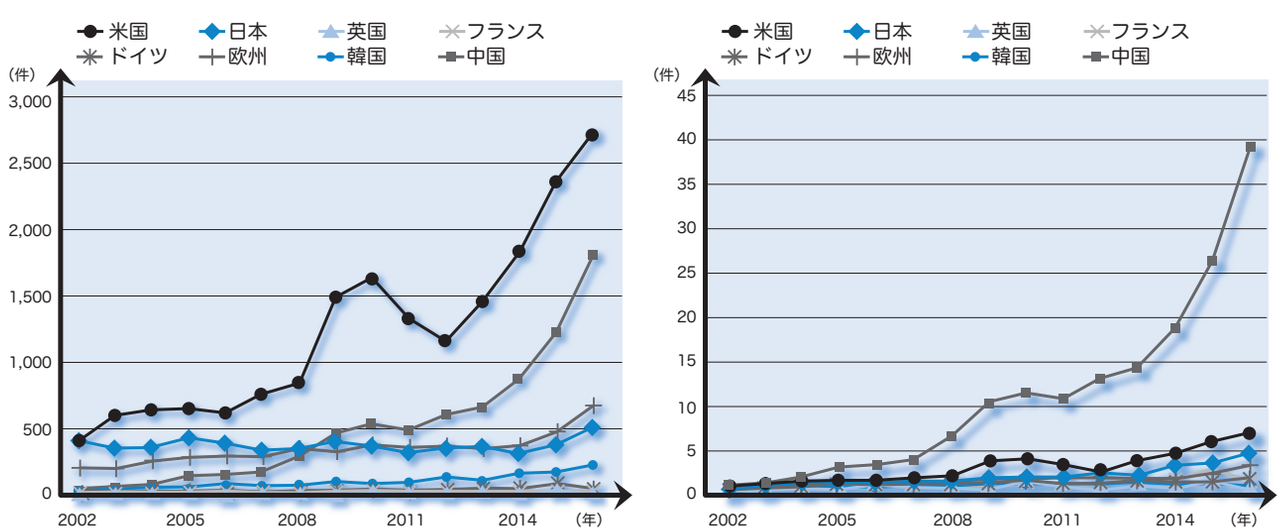
1.9.4　特許・論文の動向

1.9.4.1　特許動向

AIが様々な領域に影響を与え始めている変革期においては、広い範囲で特許権を取得できるケースは少なくない。また、AIの進展は著しく、自社技術として秘匿しても、優位性を維持できる期間は短い。そのため、国内外の主要企業等では、開発した技術に関わる特許出願を進めている状況にある。

図51に、AIに関わる特許の動向を示す。調査対象は、2002年から2016年の期間に発行された公開特許公報とし、検索にはクラリベイト・アナリティクスのThomson Innovationを使用し、国際特許分類（IPC）の分類コード「G06N」（特定の計算モデルに基づくコンピュータ・システム）を用いた[29]。

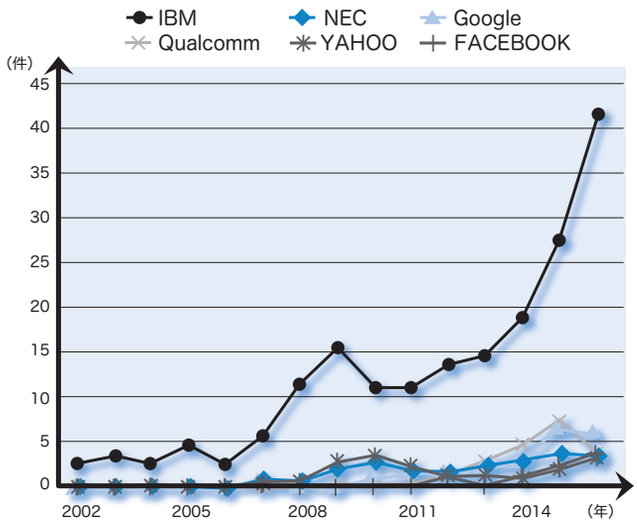
各国の特許文献数を見ると、2002年には日本が399件とトップに位置していたが、米国及び中国における発行数の伸長により、2016年では米国（2,691件）、中国（1,795件）、欧州[30]（664件）、日本（497件）と日本の特許文献数は相対的に低下しつつある。また、2002年の件数を1とした場合の2016年の文献数は、中国（39.0）、米国（6.9）、韓国（4.6）と続き、日本は1.2と横ばい傾向となっている。



■図51　各国の特許文献数（左：2002年～2016年の推移、右：2002年を1としたときの増減）

出願人	2002～2006年	2007～2011年	2012～2016年
IBM	165	550	1,166
Mivrosoft	170	334	369
Qualcomm	6	14	209
Google	3	38	202
NEC	18	82	155
SAMSUNG	23	46	111
FUJITSU	35	56	95
SONY	134	121	94
YAHOO	4	103	94
FACEBOOK	3	6	92
HP	82	63	81
Simens	24	102	72
SAP	44	67	62
D-Wave System	52	44	55
ORACLE	20	78	39

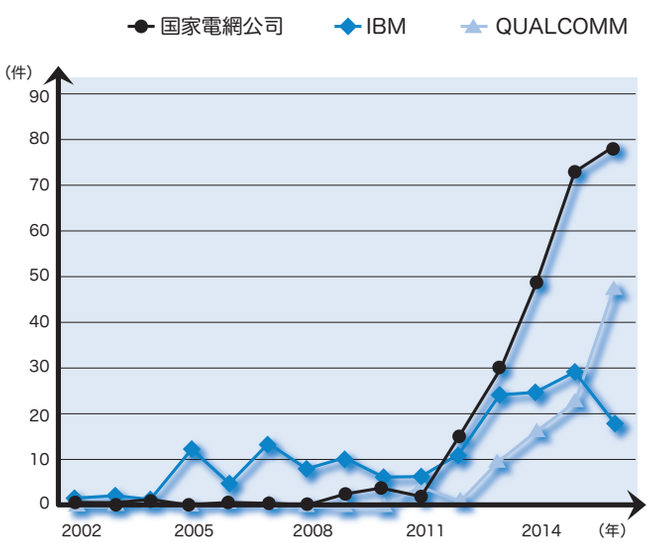
（注）各機関の上位５機関のセルをハッチングしている



■図52　米国における出願人の動向（左：上位出願人と件数、右：主な出願人の件数推移）

出願人	2002～2006年	2007～2011年	2012～2016年
国家電網公司	0	8	245
IBM	21	43	107
Qualcomm	2	5	98
Microsoft	54	42	91
西安電子科技大学	0	27	82
浙江大学	0	47	73
清華大学	3	30	61
東南大学	1	18	55
北京航空航天大学	1	61	53
SONY	51	70	53
天津大学	0	20	51
北京工業大学	2	19	50
南京航空航天大学	0	14	42
Siemens	8	24	35
Philips	13	43	16

（注）各機関の上位５機関のセルをハッチングしている



■図53　中国における出願人の動向（左：上位出願人と件数、右：主な出願人の件数推移）

次に、特許文献数の多い米国、中国、日本における国別の出願人の動向を示す。

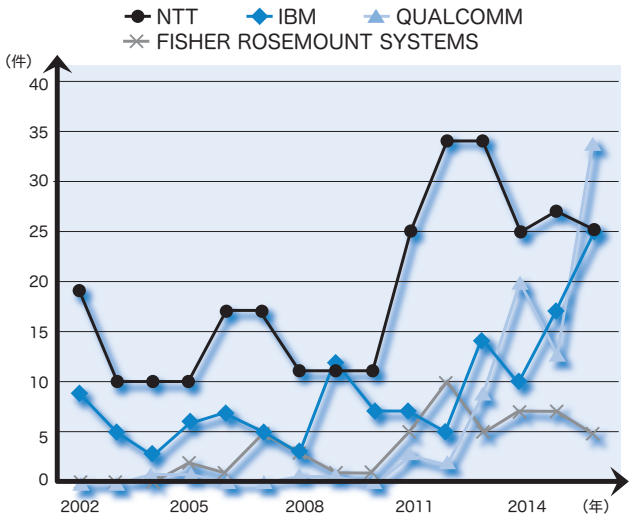
米国においては、IBMが首位であり、次いでMicrosoftが追従する状況にある。我が国企業は、2002～2011年の期間でソニーが、2012～2016年の期間でNECが上位に位置している。また、近年の件数の伸び率をみてもIBMが他社を圧倒している様子がうかがえるが、近年、Google、Qualcomm、FacebookやNECが、急速に件数を伸ばしている（**図52**）。

中国においては、中国全土へ送電・変電・配電を行う国家電網公司が首位であり、2012～2016年の

※31
論文の検索にはクラリベイト・アナリティクスのThomson Innovationを使用し、Web of science及び国際会議のプロシーディングスを対象とした。検索キーワードは“deep learning” OR “deep neural network”とし、教育学関係の検索キーワード（“education” OR “education*” OR “school” OR “student” OR “college” OR “classroom” OR “undergraduate” OR “psychology*” OR “therapy” OR “personality” OR “gesture”）を含むものを排除した。対象国は特に限定していない。

出願人	2002～2006年	2007～2011年	2012～2016年
NTT	66	75	145
FUJITSU	81	52	90
Qualcomm	2	5	78
IBM	30	34	71
SONY	104	140	68
NEC	30	60	58
TOSHIBA	71	67	47
CANON	41	29	40
HITACH	42	42	34
FISHER ROSEMOUNT	3	15	34
Microsoft	60	36	22
FUJI XEROX	39	25	17
Philips	23	40	17
HONDA	22	32	16

（注）各機関の上位５機関のセルをハッチングしている



■図54　日本における出願人の動向（左：上位出願人と件数、右：主な出願人の件数推移）

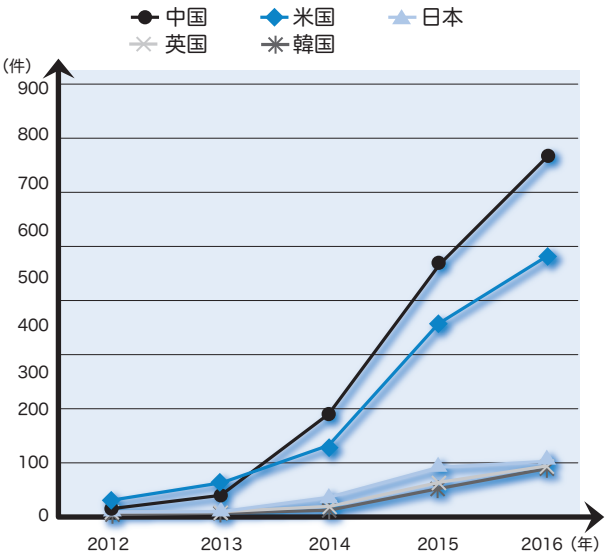
期間に245件もの件数となっている。米国のMicrosoftやIBM、Qualcommが、我が国企業では、ソニーが上位に位置している。中国の特徴として、大学からの特許出願が2007年以降増加し、公開される特許の多くを占めるようになっている（**図53**）。

日本においては、ソニーが発行総数312件で首位であり、NTT（286件）、富士通（223件）と続く。近年の傾向を見ると、IBMやQualcommなどの海外資本企業が件数を伸ばしている（**図54**）。

1.9.4.2　論文動向

図55に、ディープラーニングの研究論文の2012年～2016年の発表件数の推移を示す[31]。現在のディープラーニングの興隆の嚆矢となるのは、トロント大学のジェフリー・ヒントン氏らの2006年の研究であるが、ディープラーニングの潜在的な能力が2011年から2012年にかけて音声認識や画像認識のコンテストで認識されて以来、ディープラーニングを用いた研究は増加の一途を辿っていることが分かる（**図55**）。

国別に見ると、2013年の段階で中国が米国を抜いてトップに立っている。また、我が国も英国、韓国とほぼ同等の件数を発表している。



■図55　ディープラーニングに係る論文数の年次推移（上位5ヶ国）

プレプリントサーバである「arXiv[32]」を用いたキーワードトレンド分析[33]では、機械学習分野で2016年に前年よりも使用頻度が大きく増加したキーワードとして、残渣ネット（ResNet）（1.2.7項参照）、生成敵対ネットワーク（GAN）（1.2.5項参照）、RNN（1.2.5項参照）、ニューラル機械翻訳（1.4.4項参照）等が挙げられている。

※32
プレプリントとは、研究者が論文原稿を学術誌に発表する前に研究者コミュニティで共有するもの。プレプリントを共有する仕組みとして、コーネル大学（米国）が運営するarXivが有名である。Cornell University Library Website <https://arxiv.org/>

※33
"A Peek at Trends in Machine Learning." Medium Website <https://medium.com/@karpathy/a-peek-at-trends-in-machine-learning-ab8a1085a106>

1.10 各国の研究開発の現状

1.10.1 総論

本章では、人工知能（AI）の体系全体が、ディープラーニングの登場により、パターンの世界での学習が可能になったことにより変革され、従来難しかった記号処理の世界とパターン認識の世界の融合へ向けた端緒が見えてきた状況を記述してきた。本節では改めて、AI技術の今後の発展がどの様な順序で実現していくかについて、整理を行う。

短期的には、本章でこれまで記述してきた様に、特徴量の学習を可能としたディープラーニングにより、特に画像認識等のパターン認識の精度が人間並みに向上してきたことを受けて、パターン認識の応用開発が様々な分野で進められるものと考えられる。また、ディープラーニングと、強化学習の仕組みを組み合わせれば、ロボットの動作の学習への応用が可能である。用途をある程度限定した目的であれば、比較的短期に利用されていくものと考えられる。

ディープラーニングによるAIの適用範囲の拡大を目指すには、論理的な推論やプログラムの様な記号的処理をニューラルネットワークで実現することは避けて通れない。理解可能なニューラルネットワークを目指すという意味でも、パターン認識と記号的処理をどの様に融合すれば良いかという問題は解決すべき課題である。中期的には、この課題の解決に向けて、脳の持つ様々な機能を模倣したディープラーニングを構成していく試みと、記号を実世界と関連付けるシンボルグラウンディングを段階的に進めていくことが必要である。

更に長期的には、汎用AI（Artificial General Intelligence; AGI）の実現に向けて、自己の認識状態を認識するメタ認知の機能や、他人にも自分と同じ心があることを理解し、心の動きのモデルを自分の中にも構築した上で、他人の動きとそのモデルを同期させ他人の心の状態を推測する心の理論など、現状ではほとんど実現されていないモデルが必要となると考えられる。このような研究開発においては、認知科学、言語学、脳科学、複雑系科学、計算科学等、AIと関連する周辺諸分野の知見も十分に取り入れつつ、最終的にはAGIの実現を目指していくことが必要であろう。

以下では、ディープラーニングに基づくパターン処理と、記号的処理が段階的に融合していくものと考えられる中期的な発展に関する見通しと、長期的観点から、最終的なAIの形態ともいえるAGIへ向けた取組について整理する。

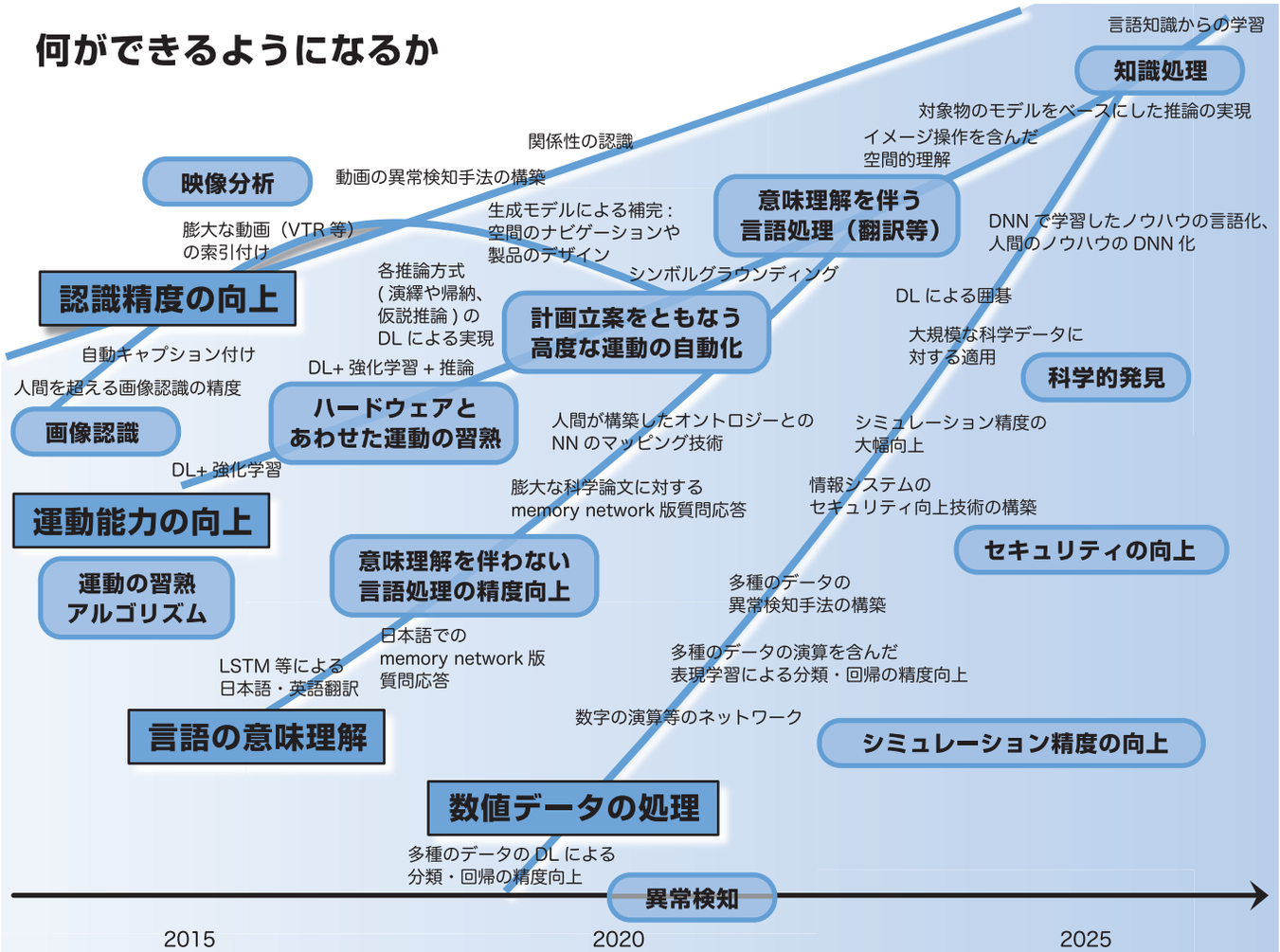
1.10.2 シンボルグラウンディングの段階的解決へ向けて

本項では、AI分野の課題の整理をした上で、今後の課題の解決の可能性を整理する。

今後、中期的には、ディープラーニングによるパターン認識と、記号的処理が融合していくものと考えられるが、具体的にどのような課題をどのような順番で実現していくかについては、必要とされる記号の意味がどの程度深いものであるかによって決まるものと考えられる。

短期的に進むと考えられる画像認識に代表されるパターン認識と、ロボットの動きの学習などは、さほど記号の深い意味に踏み込まずに処理できる領域であり、研究開発と社会実装が進むものと考えられる。その後に関しては、記号の意味を実世界の事物へ関連付けるシンボルグラウンディングというハードルを乗り越えることが必要となる。

シンボルグラウンディング問題は、AI分野で古くから課題とされてきた問題の一つである。冒頭～1.3節で述べたように、シンボルグラウンディング問題を解く鍵は、身体性の実現にある。身体性とは、メタ認知や内部的な動機に基づいて、行動（環境への働きかけ）の計画と行動の結果の予測を行い、実



■図56　今後のAIの発展の方向性[1]

行後にその影響の評価を行い、内部のモデルに還元するという認知のループを行う枠組みであり、そのループのイタレーションの中で、記号が自発的に分節されるのが記号創発の仕組みであった。

したがって、外界の事物のパターン、そのパターンの時間的発展を予測するモデルを、内部モデルとして知能の中に持つことが必要となる。外界の事物のパターンを内部モデル化することはディープラーニングで実現されたことであり、その時間的発展を予測することは生成モデルにより実現されたことである。そして、ディープラーニングが登場してから数年経った現在、シンボルグラウンディングに必要であった機能が、現在手元にあるということである。

その結果、今後のAIの技術の発展は、おおよそ次のように進むと考えられる（**図56**）。まず、既に精度の向上が進んでいる画像処理等のパターンの認識技術と、ロボティクスのハードウェア技術の進化と強化学習に基づく制御の高度化技術が融合。その融合の中で生成モデルが活用されることで、身体性の枠組みがある程度実現し、部分的にかもしれないが、シンボルグラウンディングが実現し、記号に対応したパターンとしての表現の高度化が可能となる。そこで獲得される表現を基に、自然言語処理の精度の向上も実現するものと考えられる。最終的には、シンボルグラウンディングにより獲得されたパター

※1
松尾豊「人工知能に関する技術動向と産業分野への利用可能性」(第2回　経済産業省　産業構造審議会　新産業構造部会　配布資料5)
経済産業省ウェブサイト <http://www.meti.go.jp/committee/sankoushin/shin_sangyoukouzou/pdf/002_05_00.pdf>

ンの世界と、ビッグデータに基づく記号処理の世界が、完全に近い形で融合するものと考えられる。

実世界へのシンボルグラウンディングではないが、パターンの処理と記号的処理の融合へ向けた萌芽的な研究は進められている。以下に脳の短期記憶に相当する機能をディープラーニングに取り込んだメモリネットワークによる推論の研究と、より原理的にチューリングマシンをディープラーニングで実現したNTM（Neural Turing Machine）を紹介する。

リカレントニューラルネットワーク（RNN）は、原理的にはノイマン型コンピュータで計算可能な計算を全て模倣できることが1995年には知られていた。しかし現実には、記号的処理をニューラルネットワークで実現することは、実現には時間がかかるものと考えられてきた[2]。

ところが、2015年、Facebook AI Research（米国）とGoogle（米国）からこの種の問題にチャレンジする研究が相次いで発表された[3]。Facebook AI Researchのジェイソン・ワトソン（Jason Watson）氏らによるメモリネットワークは、ディープラーニングのネットワークに、明示的に読み書き可能なメモリ機能を組み込むことで、自然言語による推論の課題の精度をLSTM（Long Short-Term Memory）等に比べて大幅に向上させた（**図57**）。

ワトソン氏らは、人工的に生成した様々なタイプの質問と回答を教師データとして学習させ、いくつかのタイプの質問に対し、100%に近い推論精度が得られること示した。

図の例において、「Where is the ring？」の問いに答えるには、指輪（ring）の動きに関し、人が指輪を持った状態で移動すると、それに伴って指輪も移動するという概念が獲得されている必要がある。このような技術は、質問応答システムや対話インターフェース等、現実的な応用先も多く、ワトソン氏らの研究をきっかけとして多くの改良モデルが提出されている。また、ニュース原稿に関する質問応答など一般的な内容に関しても約80%弱程度まで正解を出せるようになっている。

Bilbo travelled to the cave. Gollum dropped the ring there. Bilbo took the ring.
Bilbo went back to the Shire. Bilbo left the ring there. Frodo got the ring.
Frodo journeyed to Mount-Doom. Frodo dropped the ring there. Sauron died.
Frodo went back to the Shire. Bilbo travelled to the Grey-havens. The End.
Where is the ring? A: Mount-Doom
Where is Bilbo now? A: Grey-havens
Where is Frodo now? A: Shire

■**図57　メモリネットワークにより可能になった推論の例**

Googleのアレックス・グレイヴス（Alex Graves）氏らが発表したNTMも、ディープラーニングにメモリ機能を明示的に組み合わせたモデルである。チューリングマシンとは、アラン・チューリング氏が定式化した現在のノイマン型計算機の理論的なモデルであり、演算器とメモリとなるテープ、テープへの読み書きが可能なヘッドから構成されるものである。グレイブス氏らは、このメモリと読み書きのヘッドをディープラーニングに組み込み、全てが逆誤差伝播法で学習可能な様に定式化を行った。NTMにより、入力の出力へのコピー、ソーティングや連想記憶など、比較的簡易なアルゴリズムがニューラルネットワークで実現できることを示した。

※2
ただし、リカレントニューラルネットワーク（Recurrent Neural Network; RNN）は、原理的にはノイマン型コンピュータで計算可能な計算を全て模倣できることが証明されている。

※3
Facebook AI ResearchのメモリネットワークのarXivへの投稿の5日後に、GoogleのNeural Turing Machineが同じくarXivへ投稿された。

1.10.3 汎用AIに向けて

1.10.3.1 目標としての汎用AI

更に長期的には、上記の様なメモリ機能だけでなく、人間の知能が持つ機能を広汎に持つAIの実現へ向けた研究が必要となる。現在実用化されている対象とするタスク（例えば、囲碁、自動運転など）に特化したAIは特化型AIと呼ばれ、対象とするタスクに応じた事前知識が潤沢に組み込まれている。これに対し、人間のように十分に広範な適用範囲と強力な汎化能力を持ち、多種多様な問題の解決することが可能なAIをAGIと呼んでいる。

AIの登場以来、人間のような知性の実現は究極目標であったが、20世紀の間はあまりにも遠いものと考えられていた。しかし最近のディープラーニング技術の進展などを契機としてその目標は到達可能なものとなってきた。こうした背景において、技術的に定義し得る人間レベルの知能の研究目標として、汎用性が着目されている。AGIが備えるべき特徴としては、経験からの学習を通じて様々な問題に対する多角的な解決能力を獲得できること、人間と同程度に多種多様な知的能力を発揮できること、等が挙げられる。

逆に「AGIは何ではないか」を指摘するなら、以下のようにいえるだろう。

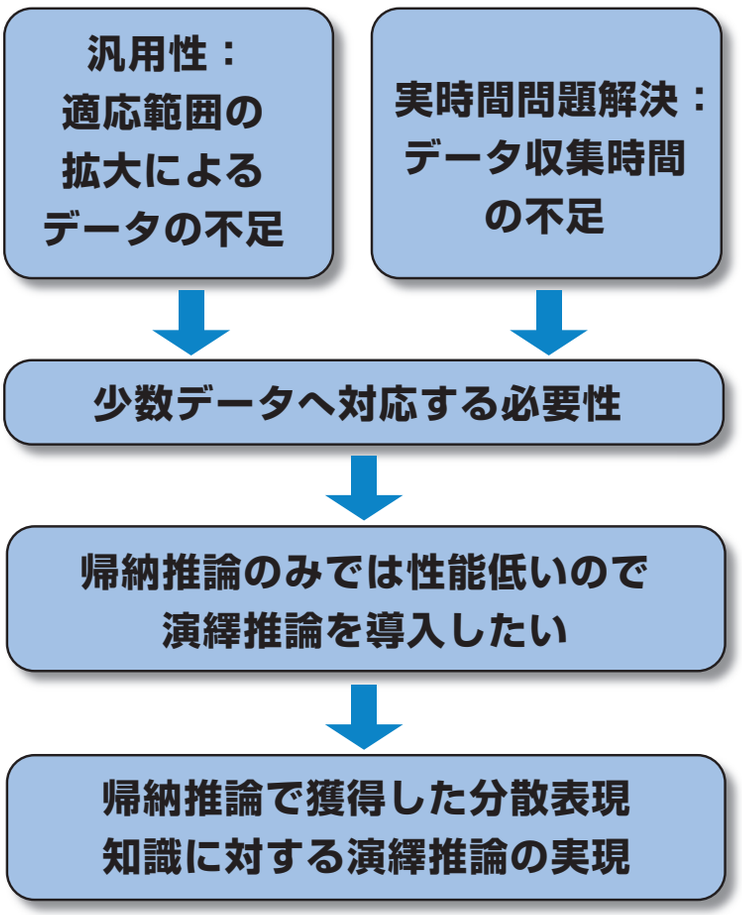
- 単に特化型AIの寄せ集めではない
- 最初から何でもできる知能ではない
- タブラ・ラサ（白紙）から学習するのではない
- 意識の有無は考慮しない（評価が困難）

AGIは任意の人間の知的活動を代替し効率化する技術であることから、経済成長及び生活レベルの向上に大きく資する。そして今世紀半ば頃には、人間個人の知能全般を凌駕し、我々の生活を支える生産者とその管理者の役割をおおむね担う技術と予測されている。一度人間レベルのAIが作られれば、それ自身を使って新たなAIを設計・製造（再帰的自己改修）することが可能になる。その際には生物の進化とは比較にならない速度で自己再帰的に発展し、人類がこれまで行ってきたペースに比べると格段に早い速度で、知を蓄積することができる。この大きな変化は、しばしば「技術的特異点」（Technological Singularity）と呼ばれる。

AGIを目指す主要なアプローチとして、外界から情報を取り入れて、何らかの意味で適切な意思決定や行動、制御（若しくはその支援）を行うための認知アーキテクチャの研究がある。認知アーキテクチャでは、人間の総合的な認知機能をモデル化しており、AIの創成期より様々なモデルが研究・開発されてきている。

例えば、カーネギーメロン大学のジョン・アンダーソン（John Anderson）氏らによって作られた「ACT-R」（Adaptive Control of Thought-Rational）[1]では、人間の認知機能を外界のオブジェクトを認識するための視覚モジュール、目標と意図を記録している意図モジュールなどに分解し、それらが協働することによって、人間と同じような機能を実現する。従来は、心理実験などを通して、認知アーキテクチャの妥当性が測られることが多かったが、近年の脳計測技術の発展により、脳を直接分析することで認知アーキテクチャを構成しようとする、生物からヒントを得た認知アーキテクチャ（BICA）も注目を浴びており、代表的なアーキテクチャとして「LEABRA」、「Micro PSI」、「LIDA」等がある[4]。

※4
"The Biologically Inspired Cognitive Architectures Society."
BICA Society Website <http://bicasociety.org/>



■図58　先鋭化する汎用性をめぐる課題

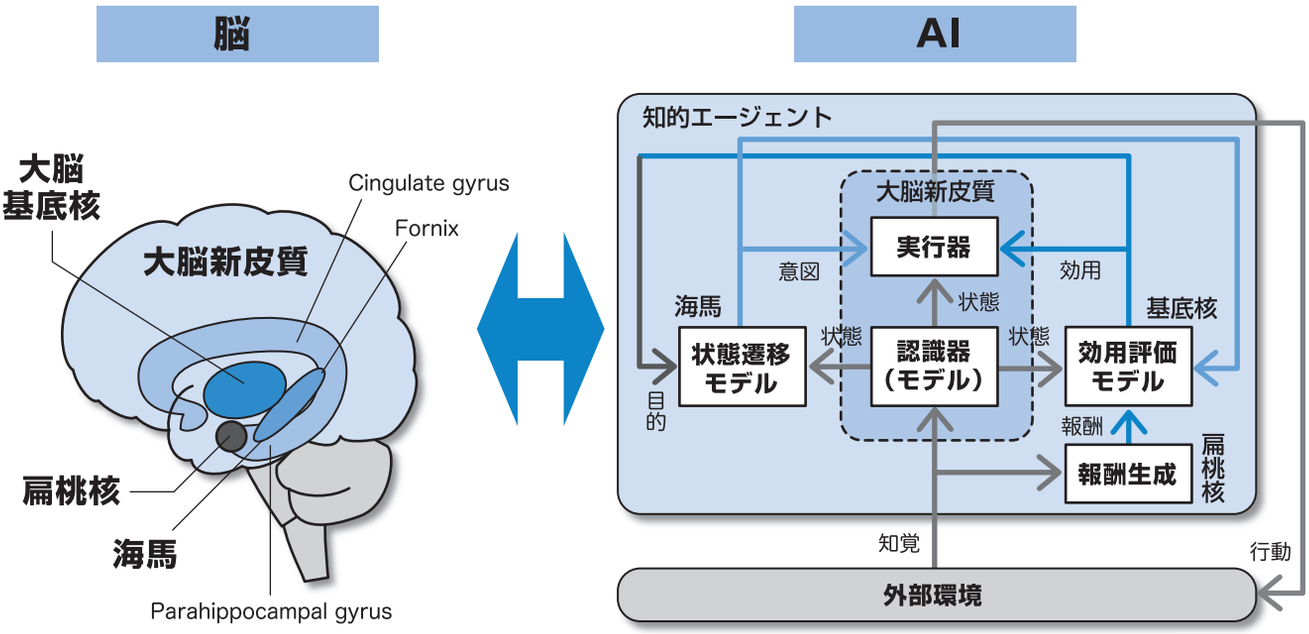
十分にデータを得られるタスクの範囲内であれば、応用価値のある人間並みの性能を持つ学習をすることが可能になった。そこで、最近のAGIの国際会議などにおいて議論されるAIの基本課題は個別のタスクの解決よりも、主に以下のような側面が着目されている。

- 汎用性
- 現実的な時間内での問題解決
- 少数データへの対応
- 演繹推論

実はこの四つの課題は**図58**に示すように深く関係する。知能の汎用化を目指してタスクの適用範囲を広げていけば、しばしばデータ不足が生ずる。ここで現時的な時間内に問題解決を行おうとすれば、十分なデータを収集するほどの時間的余裕がない。こうしてデータが不足は解消できない状況では帰納推論のみでは良い性能は得られない。そこで、演繹推論を導入することで、データが少ない新たな状況にも対応する必要がある。こうした側面は、以前より繰り返しAI分野で議論されてきたが、ディープラーニングの進展を背景に、AGIという研究の文脈において鮮明化している。

1.10.3.2全脳アーキテクチャ・アプローチ

現存するAGIは脳以外には存在しないため、脳の構造と機能を模倣した計算モデルを構築していくことがAGIの研究を進める有力な手段となる。全脳アーキテクチャ（WBA）・アプローチは、「脳全体のアー



■図59　全脳アーキテクチャ・アプローチ

キテクチャに学ぶことで人の様なAGIを創る」という工学的なアプローチである。

こうした脳型AIの研究に資する最新の技術が、ディープラーニングと脳のイメージングである。ディープラーニングは、脳において汎用性を担う大脳新皮質を模倣している計算モデルとみなせる。神経科学においては、光遺伝学等の進歩により、多くのニューロン活動を同時計測し制御できるようになり、脳全体の静的なネットワーク構造や、局所的な詳細なネットワーク構造がコネクトーム研究で明らかにされつつある[2]。こうしてニューロン数個のミクロな振る舞いと、行動につながる脳全体のマクロな振る舞いとが関連した理解が進んだことで脳をAIとして理解し構築できるようになった。

NPO法人全脳アーキテクチャ・イニシアティブ（WBAI）は、2030年までに公益的な立場から上記アプローチによるAGIの完成を促進することを目的として2015年に創設され、全脳アーキテクチャ（WBA）を進めるための技術基盤として学習環境シミュレータLIS（Life in Silico）等を構築するとともに、ハッカソンの実施や、さらに勉強会を通じての人材育成、啓蒙活動などをオープンにすすめている。

AI技術では必ずしも人間の脳を参考にする必要はなく、脳の模倣は時として足枷にもなる。しかし、人間のレベルを凌駕するまでは、人間には設計が難しい未解決な計算機能についてのヒント（暫定モデル）が得られるメリットがある。更に理解が進みつつある脳全体に対応する認知アーキテクチャは、分散共同開発において合意しうる技術統合の足場となるため、脳に学んだAGIのオープンな開発が可能になる。

このほか、2016年には国内においても、新学術領域研究（人工知能と脳科学の対照と融合）や文部科学省（ポスト「京」萌芽的課題「全脳シミュレーションと脳型人工知能」）の開始、ソニーのCogitai（米国）への資本参加、東京大学の次世代知能科学研究センターの設置、電気通信大学のAI先端研究センターの設置、AIエンジンの開発を進めるベンチャーであるDeep Insightsの設立といったAGIに向けた動きが次第に活発化してきた。

1.10.3.3.　海外におけるAGIへの取組

AGIを構築しようとする試みは、2006年頃にベン・ゲーツェル（Ben Goerztel）氏により提唱され、近年では2015年にDeepMind（英国）、GoodAI（チェコ）、OpenAI（米国）などがAGIの開発を推進していくとの方向性を明確にしている。2016年には、ラットレベルのAGIの開発を宣言したDeep-

Mindが開発成果を続々と発表している。DeepMindは、比較的脳に学ぶ形での研究開発を進めているものと考えられる。

優秀なAI研究者のレベルの創造性を持って更に高度なAIをプログラミングできるAIが造られれば、AI自身による自己再帰的な改良が可能になる。そのレベルに達した最初のAIは、「Seed AI」と呼ばれ、その出現が技術的特異点の起点となる可能性が高い。1980年代頃までに行われたプログラミングを行うAI研究はその後収束していた。しかしDeepMindのチームが2014年10月にディープラーニングにメモリを結合したNTMを提案し、ソートアルゴリズムを自動学習しうることを示すなど、Seed AIに繋がりうる研究成果を示し、2015年にはNPI（Neural Programmer Interpreters）などの成果が公開されている。

GoodAIはAGIを開発することのみを目的に2015年に創設されたチェコの会社であり、できるだけ早く人類を助け世界を理解するためのAGIを完成することをミッションとし、AGI開発のロードマップを公開した上で、5百万ドルのチャレンジを公募している。

この分野ではAGIを育成するためのデータの生成が必要になるため、ゲーム環境を利用した学習環境の公開が進んでいる。WBAIにおいても「LIS」を開発してきたが、OpenAIは、2016年の初頭に主に強化学習のための環境として「OpenAI Gym」を提供し2016年末には「UNIVERS」という環境を公開した。この時期には、「DeepMindはDeepMind Lab」を、Facebookは「TorchCraft」を公開している。

参考文献
[1] John R. Anderson et al., “An Integrated Theory of the Mind,” Psychological Review, vol.111 No.4, pp.1036-1060.
[2] セバスチャン・スン『コネクトーム:脳の配線はどのように「わたし」をつくり出すのか』草思社.