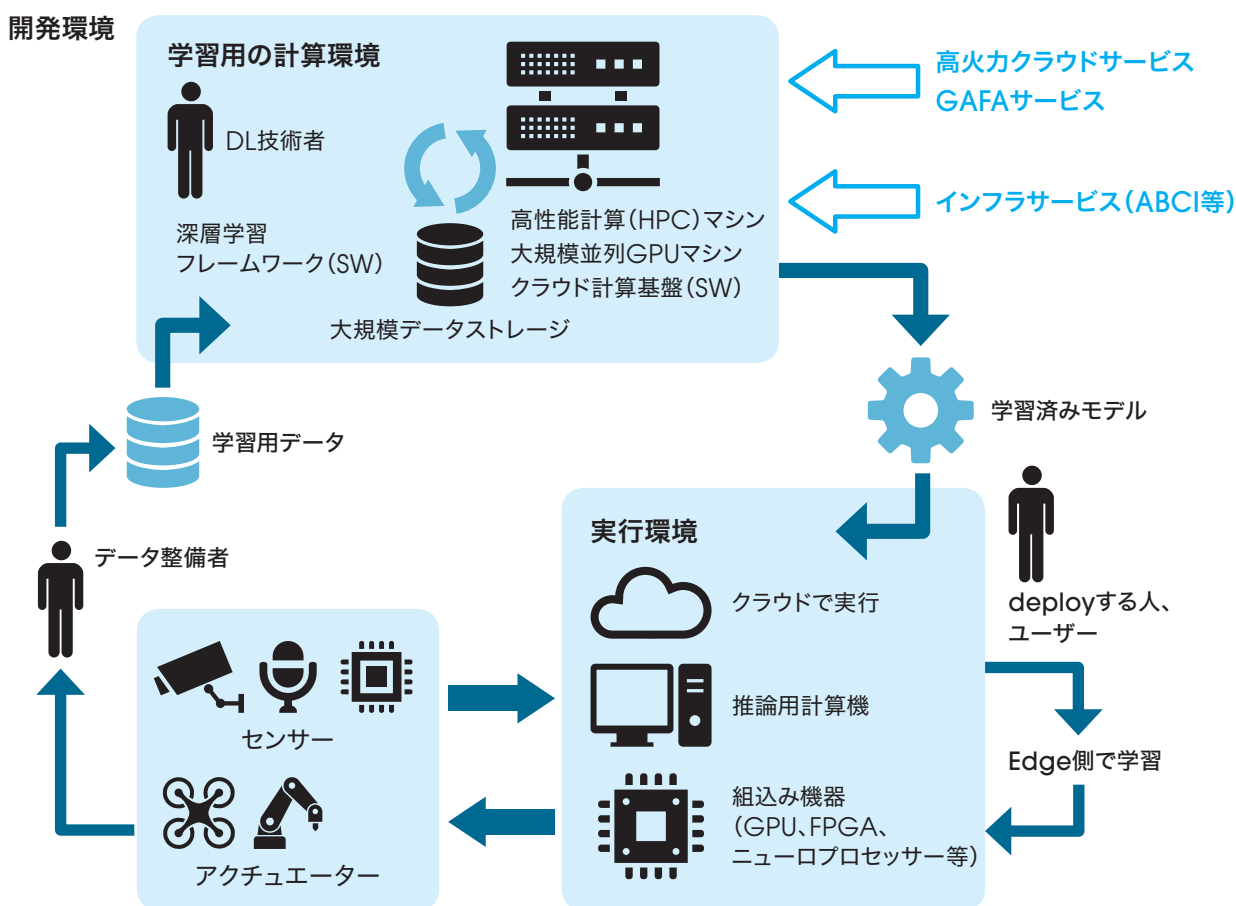


## 2.8 ▷ 開発基盤

AI、特にディープラーニング(深層学習)に代表される機械学習の利用においては、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に分けられる。学習フェーズでは、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、膨大な計算能力が要求される<sup>\*75</sup>。一方、推論フェーズでは、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ないが、データの入出力や格納、転送に関して高い能力が求められる。

以下ディープラーニングに関わる開発環境を中心に、インフラやハードウェアの開発状況について説明する。ディープラーニングの利用者は、上記①、②各々のフェーズに合わせて計算機資源や推論環境を切り替えながら開発を行う。図2-8-1は典型的な開発・利用の例である。

■図2-8-1 典型的なディープラーニングの開発・利用フロー



データ整備者は、センサーからのデータ、アクチュエーターのデータを集めて、整形やアノテーション(正解ラベルの付与)を行い「学習用データ」を作成する。ディープラーニング技術者は、「学習用データ」に対して、学習フレームワーク(SW)やクラウド計算基盤(SW)を利用して、分析しながら学習用のプログラムを作成する。プログラムが完成したら、GPU等の学習に適切なハードウェア資

\*75 例えばAlphaGo(碁)においては、戦略決定用のCNNはパラメーター数は400万、畳み込みの足し算回数には14億回を要する。3,000万局面の学習を行うのに50GPUで3週間かかるとすると、CPU単独ならば300週(6年弱)はかかる計算になる。

源を潤沢に持ち並列計算に優れた高性能計算 (High Performance Computing ; HPC) マシンを利用して学習モデルを作成する。この高性能計算機は、ローカルに整備してもよいし、クラウドサービスを活用したり、外部のインフラサービスを活用するなど複数の選択肢がある。

「学習モデル」ができると、ローカルの実行環境にモデルの適用 (deploy) を行い、センサーデータを入力した推論を行って、結果を出力したり、アクチュエーターを操作するなどのアクションを行う。実行環境では、ローカルPCの利用、クラウド計算の活用などで推論を実行するが、組み込み機器に直接モデルを転送して、FPGA (Field Programmable Gate Array) やGPU (Graphics Processing Unit) を使って推論計算なども行う。また近年ではローカルの計算機環境の性能向上により、エッジ側で学習処理ができるようになる等の学習と実行を同一の環境で実行するような利用例もある。

学習フェーズにおいては高性能計算が必要であり、ディープラーニングと相性のよいGPUを用いたいわゆる高速計算が簡単に利用できるようになったことが、近年のディープラーニングの発展につながっている。一方、HPCにおいては我が国には長い研究開発の蓄積がある。AIの研究開発や実用化に際して、HPC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。

一方、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、短レイテンシー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。

本節では、ディープラーニングで要求される演算の基本と、HPCによる並列化とその課題について紹介し (2.8.1 (1))、GPUに代表されるディープラーニング向けのプロセッサ技術の動向 (2.8.1 (2))、学習時での高性能計算機インフラストラクチャーと、国レベルで本格的に実施する計算のための大規模環境の整備の動向として、産業技術総合研究所 (AIST) の「AI Bridging Cloud Infrastructure (ABCI)」を紹介する (2.8.2)。そして、推論用、特にエッジ向けのFPGA、システムオンチップ (SoC) などの最新動向を紹介し (2.8.3)、最後に新世代のAIインフラストラクチャーとして、従来のノイマン型コンピューターと異なる計算原理である非ノイマン型コンピューターや、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、量子計算等の発展について紹介する (2.8.4)。

## 2.8.1 基本原理

### (1) 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。

#### 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すべ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(パラメーター・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ー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ーのスライド幅(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：

パラメーター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。

### ● WinogradのMinimal filtering algorithm：

Minimal filtering algorithm[1][2]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ー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り、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ーサイズが必要なFFT (Fast Fourier Transform)と比較してフィルターサイズが小さい場合に特に有効である。

### ● FFTを用いる方法：

2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(Convolutional Neural Network; 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：

米国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<sup>\*76</sup>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、ユーザーが選択できるようになっている。これにより、ユーザーが(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。

---

\*76 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(Depth Neural Network; DNN)用のライブラリ。

## 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(Deep Neural Network; DNN)のパラメーターを最適化する最適化問題に帰着される。パラメーター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。(図2-8-2参照)

SGDではパラメーター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回の更新当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当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ーに対するコスト関数の勾配を計算する手法である。

■ 図2-8-2 DNNのパラメーターの最適化に用いられる確率的勾配降下法(SGD)

\*DNNのパラメーター $\theta$ を最適化する最適化問題

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

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)、 $E_{\theta}$ 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。

\*パラメーターの最適化に用いられる確率的勾配降下法SGDの反復式

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

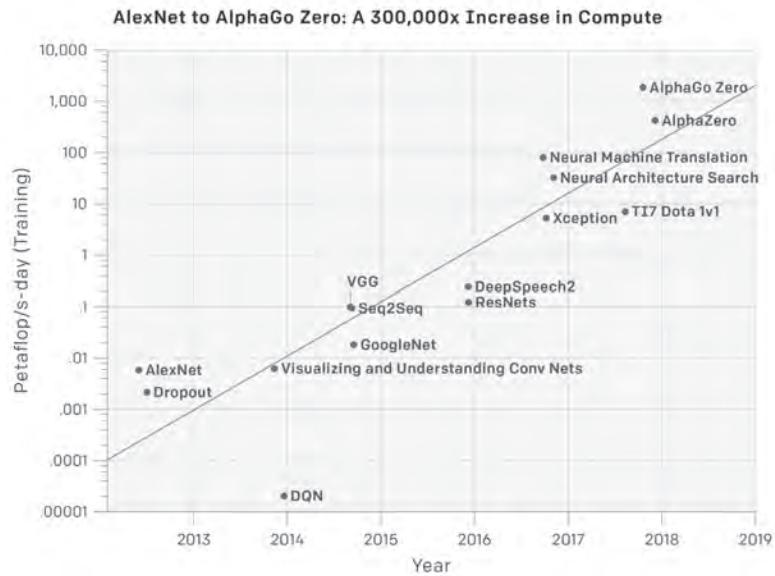
ただし $M^{(t)}$ はステップ目でデータセットからランダムに選択された部分集合(ミニバッチ)、 $\eta > 0$ は学習係数である。

## ③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用の計算需要の増加は、3.5カ月ごとに倍増している(図2-8-3)ともいわれている。また、ビッグデータ処理能力の大幅な増加も予想され、何ペタバイトにも上るデータを扱う必要がある。

■ 図2-8-3 急増するディープラーニングの計算需要



出典: "AI and Compute" OpenAI blog

すでにGPUの採用が進んでいるが、総合的に鑑みればハードウェアアーキテクチャーとして最適であるものは、現代の超並列アーキテクチャー型のスーパーコンピューターである。世界の上位ランクのトップスーパーコンピューターは、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、さらにテラバイト/秒の超高性能のI/Oを備えており、このようなシステムを用いた計算はHPCと呼ばれる。

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にあるが、ディープラーニングでは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ー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り、これらへの対策も検討されている。例えば、SqueezeNet<sup>\*77</sup>のようにネットワーク中のフィルターやそこに通すチャンネル数(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などである。

## 2) ディープラーニングの大規模並列化とその課題

スーパーコンピューター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。HPCの分野では並列化・高性能化の研究や実装が長年行われてきた分野であるが、いくつか深層学習に特化するべき部分がある。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。

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: feed forward計算で生じる活性化等)のことを表し、「モデル」とはDNNそのものを表す。

\*77 "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size", F.N. Iandola, et.al <<https://arxiv.org/abs/1602.07360>>

- データ並列：

複数の計算機が同一のDNNのパラメーターを持ち、異なるデータについて並列に計算を行う手法。DNNのパラメーター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。高頻度な通信が必要となるが、計算機同士でDNNのパラメーターを重複して持つ必要がないため、パラメーター数が大きく単一の計算機のメ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ー数が小さいため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ー数が大きいいため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ー）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2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は、GoogleのDist Belief、MicrosoftのProject Adamが採用した「パラメーターサーバー」や、BaiduのDeep ImageやDeep Speech、MicrosoftのCNTKで採用された「集団通信による手法」などがある。

### 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確率的勾配降下 (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。さらにデータ並列、モデル並列、非同期学習 (パラメーターサーバーを含む) 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

- モデル並列：

レイヤーごとに全対全通信が必要となるので、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ー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し、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うえで重要なことは、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。性能の指標としても、従来の計算性能 (1 エポック<sup>※78</sup>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 (十分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。

---

※78 データセット全体を使った1回分の学習のこと。

## (2) ディープラーニング向けプロセッサ技術の動向

ディープラーニングの学習・推論に使われる計算のうち代表的なものとして、フィードフォワード計算とバックプロパゲーション計算がある(2.8.1項参照)。これらの計算においては、大量のデータ列に対する積和演算などに要する計算量が支配的であり、ディープラーニング向けプロセッサはこうした計算を高速実行することを意図した設計がなされている。

### ①ディープラーニング向けGPU

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・パラメーターに対して同一の計算を適用することが多く、グラフィックス処理用のプロセッサとして開発されたGPU(Graphic Processing Unit)はこうした計算の実現に特化した設計となっている。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(Single Instruction Multiple Data)並列計算する機能を備えている。General-Purpose computing on GPU(GPGPU)は、かつてのFPUと同様、こうしたGPUを汎用の拡張計算機能として利用するためのプログラミング環境であり、NVIDIAのCUDA(Compute Unified Device Architecture)、AMD(米国)のAMD Stream、クロノス・グループによる標準規格OpenCL(Open Computing Language)などが広く知られている。GPU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、条件分岐が混ざらない低精度の浮動小数点演算でよい性能を上げる。前述のとおり、低精度の積和演算は、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。

具体的には、数千個の演算コア(CUDAコア、またはストリーミングプロセッサなどとも呼ばれる)と広帯域メモリをパッケージ化することで、SIMD的な処理の性能を追求する一方、仮想メモリの実現や複雑な分岐処理などの性能は犠牲にしている。これらの工夫により、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。

ディープラーニング向けのGPUとしては、すでにデファクトとしての地位を確立したNVIDIAのTeslaのほか、AMDのRadeon Instinctがある(表2-8-1)。

■表2-8-1 ディープラーニング向けGPUのスペック

メーカー	NVIDIA		AMD
製品名	Tesla P100 for NVLink	Tesla V100 for NVLink	Radeon Instinct MI25
FP16 / FP32 混合		125 TFLOPS	
FP16	21.2 TFLOPS		24.6 TFLOPS
FP32	10.6 TFLOPS	15.7 TFLOPS	12.3 TFLOPS
FP64	5.2 TFLOPS	7.8 TFLOPS	0.8 TFLOPS
HBM2 容量	16 GB	16 GB / 32 GB	16 GB
HBM2 帯域	732 GB/s	900 GB/s	484 GB/s
演算コア数	3,584	5,120	4,096
マイクロアーキテクチャ	Pascal	Volta	VEGA

出典:各種公開資料より作成

NVIDIAは、ディープラーニング向けの大型のGPUとして2016年4月にTesla P100、2017年5月にTesla V100をそれぞれ発表した。Tesla P100(Pascalマイクロアーキテクチャー採用)では、従来の演算ユニットであるCUDAコアをディープラーニング用途に利用する。つまり、FP32(単精度、32ビット)の積和演算を学習に、FP16(半精度、16ビット)の積和演算を推論に利用することを想定している。Tesla V100(Voltaマイクロアーキテクチャー採用)では、従来のCUDAコアの他に、ディープラーニングの学習と推論の両方に対応したTensorコアと呼ばれる新しい積和演算ユニットを640基搭載する。Tensorコアは、4x4行列A、B、Cを入力として $A \times B + C$ のテンソル計算を1サイクルで行う(A、BはFP16、C及び計算結果はFP32となる混合精度)。Tensorコアは、125 TFLOPSと従来のCUDAコアよりも一桁近く高いピーク性能を持ち、このTensorコアを有効活用したアプリケーションの開発が今後の課題となる。

NVIDIA Tesla P100/V100のもう一つの特徴は、GPU間的高速通信を実現するNVLinkと呼ばれる高速インターフェースを装備していることである。それまでのGPUでは、GPU間の転送速度はPCIバスに律速され、32GB/secが上限であったのに対し、Tesla P100では双方向160GB/sec、V100で双方向300GB/secの転送が行えるようになった。NVLinkは、ディープラーニングの計算処理を、計算サーバーに搭載した複数のGPUで並列化して高速実行可能にした。また、NVIDIAはNVLinkで結合されたTesla P100/V100を8基搭載したディープラーニング専用サーバーDGX-1、Tesla V100を16基搭載したDGX-2を開発し、多くの人工知能技術の研究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。

一方、NVIDIAは、リアルタイム推論用途に特化したGPUもシリーズ化している。2016年9月にはPascalマイクロアーキテクチャーをベースとしたTesla P4、2018年10月にはVoltaマイクロアーキテクチャーの後継となるTuringマイクロアーキテクチャーをベースとしたTesla T4をそれぞれ発表している。Turingマイクロアーキテクチャーの特長の一つはTensorコアを拡張した多精度Tensorコアの採用であり、INT4(整数、4ビット)、INT8(整数、8ビット)の積和演算の高速化が図られている。これにより、INT8精度ではTesla T4はP4の6倍近い性能を示す。これらの製品は、Tesla P100/V100の4分の1、75Wの低消費電力で駆動するため、データセンターにおけるスケールアウト構成にも適している。

AMDは、ディープラーニング向けGPUとしては後発だが、2017年にRadeon VegaシリーズのGPUとしてRadeon Instinct GPUシリーズを発表した。このうち、学習向けの製品であるMI25はFP32で12.3 TFLOPS、FP16で24.6 TFLOPSと、TeslaシリーズのCUDAコア性能と同等の性能となっている。

開発環境の点では、従来GPUで動作するプログラムを作成するには、CUDAやAMD Stream、OpenCLなどGPUに特化した開発環境を用いる必要があった。しかし、近年ではGPUに最適化された各種ライブラリの充実や、既存プログラムに「ディレクティブ」と呼ばれる記述を付加するだけでGPU用のコード生成を行うOpenACC(Open Accelerators)の導入により、習得コストの低いプログラミング言語・方式で、GPUが実用可能となっている。

## ②その他ディープラーニング向けプロセッサ

Intel(米国)は、HPC向けに高性能メニーコアプロセッサであるXeon Phi Processorシリーズを数年来開発している。2016年に発売された72X0(Knights Landing、KNL)プロセッサは、米国エネルギー省Sandia国立研究所のTrinityや同Lawrence Berkeley国立研究所



のCori、我が国では東京大学OakForest-PACSなどに採用されている。2017年に発売された72X5 (Knights Mill、KNM) プロセッサは、KNLをベースに作られたディープラーニング向け高性能メニーコアプロセッサである。KNMは、KNLのFP64 (倍精度、64ビット) の演算パイプラインを半分の1本、FP32の演算パイプラインを2倍の4本、さらにVNNIパイプラインを4本新設したものである。VNNIはINT16 (16ビット整数) とINT32 (32ビット整数) の混合精度の演算を行うもので、Tensorコアなどがサポートする仮数部を表現するビット数に制限があるFP16/FP32の混合精度演算に比べて、より高い学習性能が得られるとIntelは主張している。

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ている。その第1弾であるLake Crestは2017年に特定顧客向けにサンプル出荷され、ソフトウェアスタックを含めた評価が行われているという。2019年末までにはその後継として開発しているSpring CrestをNervana NNP-L1000という製品名で製品化するとしている。

日本では富士通が、ディープラーニング専用プロセッサであるDeep Learning Unit (DLU) を開発している。DLUは、ディープラーニングの計算に適した演算精度DL-INTを導入すること、ディープラーニングに必要な機能に絞ってマイクロアーキテクチャーを簡素化すること、複数のDLUを接続するためのインターコネクトとしてTofuを応用したネットワークを採用することなどが特徴であるとされる。第1世代のDLUについては、2019年3月までのリリースを目標としているとのこと。

Googleは、自社サービスで利用することを目的として2015年にTensor Processing Unit (TPU) v1、2017年にTPU v2、2018年にTPU v3を開発し、導入を進めている。TPU v1はINT8 (8ビット整数) の積和演算器をシストリックアレイ状に256×256基搭載した推論専用のプロセッサであり、検索のランキング、音声認識、テキストやスピーチの翻訳、写真の検索などのWebサービスに利用されている。

TPU v2は、v1で課題となっていた外部メモリからのデータ呼び出しのバンド幅を改善するため16GBの広帯域メモリを搭載した。また、Google独自のBrain Floating Formatと呼ばれる16ビット浮動小数点数形式 (bfloat16) による128×128シストリックアレイを2基搭載しており、推論だけでなく学習も実行できるプロセッサとなっている。bfloat16は、同じ16ビットの浮動小数点数形式FP16が指数部に5ビット、仮数部に10ビット割り当てているのに対し、指数部に8ビット、仮数部に7ビット割り当てており、FP16に比べると数値の精度は低いが、数値の範囲は広いという設計となっている。TPU v2の性能は単体で180 TFLOPS、TPU Podと呼ばれる4ラック構成、64基のTPU v2を搭載した機械学習向けスーパーコンピューターで11.5 PFLOPSに達する。2018年に発表されたTPU v3の詳細は明らかになっていないが、液冷システムを導入することで実装密度を2倍とすることを可能にした結果、8ラック構成、256基のTPU v3を搭載したTPU Podで100 PFLOPSの性能を実現するという。

## 2.8.2 クラウド側基盤 (ディープラーニング向け計算インフラストラクチャーの動向、ABCI)

### (1) ディープラーニング向け計算インフラストラクチャーの動向

ディープラーニング向け計算インフラストラクチャーの整備は、民間のメジャーなクラウドベンダーが先行している。特に米国や中国において、GPUや高速ネットワークなどのスーパーコンピューター由来の技術を取り入れたAI向けのクラウド型計算インフラストラクチャーを急速に拡充している。一方、公的なスーパーコンピューター側も従来のシミュレーション利用一辺倒から、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で構成されたマシンはデータ解析やAI利用への対応が進んでいるが、専用プロセッサで構成されたマシンではハードルがまだまだ高い。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においては、1,000万プロセッサコア数を超える世界のトップランクの公的なスーパーコンピュータがシミュレーション研究等に供されているのに対して、AI向けにおいては、民間のインフラストラクチャーのほうが豊富であり、広範に利用されている。しかしながら、後者の利用コストは非常に高く、裾野の広い研究の阻害要因となっている。2017年3月現在、Google、AmazonなどのGPUの計算ノードのTFLOPS当たりの単価は月額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。つまり、現在のAI研究においては、基礎研究の段階でさえ民間の計算インフラストラクチャーを使う必要があり、競争力のある研究を行おうとすれば数千万円単位のコストがかかる状況にある。その一方で、欧米の公的スーパーコンピュータ全般の利用権や、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、審査ベースで割り当てられており、基本的には無料で利用できる状況にある海外と比較すると、AI研究の置かれている研究環境は極めて貧弱と言うほかない。特に大規模な計算インフラストラクチャーを自前で保有し、競争力のある研究が行えているGoogleやBaiduなどの附属研究所は、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。計算インフラストラクチャーのアクセシビリティが彼我の差を生んでいると言っても過言ではない。

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ーの整備も行われ始めている。特に、我が国では、表2-8-2に示すとおり、国立のAIの3つのセンターやその協力機関などにおいて、AI向けの公的な計算インフラストラクチャーの整備が急ピッチで進んでいる。

■表2-8-2 我が国のAI専用・AI向け公的インフラストラクチャー

導入年月	システム名	システム概要	理論性能値 (深層学習向け精度)	研究機関
2017年4月	AAIC (AIST AI Cloud)	NVIDIA Pascal P100 × 8 GPUサーバー × 50台	8.4 PFLOPS	産業技術総合研究所 AI 研究センター (AIST-AIRC)
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 サーバー) × 25台	4 PFLOPS以上	理化学研究所 革新知能統合研究センター (Riken AIP)
2017年8月	TSUBAME3.0+2.5+KFC (HPC共用)	NVIDIA Pascal P100 x 2160 + K20X × 4080 × K80 × 168	65.8 PFLOPS	東京工業大学 学術国際情報センター (Tokyo Tech. GSIC)
2018年6月	ABCI (AI 橋渡しクラウド)	NVIDIA Tesla V100 × 4GPU サーバー × 1088 基	550 PFLOPS	産業技術総合研究所
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2-8-2に掲載した以外では、公的なスーパーコンピュータセンターにおいては、新規の設計・調達に2カ年近い時間がかかり、かつ既存のユーザーベースをないがしろにすることも困難であるため、HPCとデータサイエンスが両立するような戦略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) を例外として、即時の対応は困難である。また、民間のクラウドベンダーにおいても、機械学習に供する大規模な計算インフラストラクチャーを実現するに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り、電源供給・冷却・ネットワークの高密度実装、さらには運用面における対応が大変困難である。例えば、AI・

ビッグデータ指向のスーパーコンピューターである東京工業大学TSUBAME3.0では、1ラック当たりの熱密度が最大61kW、国立研究開発法人 産業技術総合研究所(以降、産総研という)のAI橋渡しクラウドでは最大67.33kWであるが、これは一般IDCのラック当たりの熱密度3~6kWと比較すると10~20倍にも相当し、このような計算インフラストラクチャーを通常のIDCに入れるのは困難である。

産総研のAI橋渡しクラウド(AI Bridging Cloud Infrastructure; ABCI)のミッションの一つは、こうした事態を打開することにある。ABCIの目標としては、①我が国に米国の民間クラウドベンダーにも匹敵するような550PFLOPS以上の、ディープラーニングを中心とした人工知能処理向けの大規模かつ超省電力な計算インフラストラクチャーを整備し、官民のAI研究者に提供すること、②そのような計算インフラストラクチャーの構築・運用のシステム技術を研究開発すること、③その計算インフラストラクチャー技術を迅速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。東京大学柏II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取組みを進めている。

米国でも、各研究機関と民間クラウドベンダーとの協業も進む一方、国立科学財団(National Science Foundation; NSF)、エネルギー省(Department of Energy; DoE)などがAI・ビッグデータ指向のインフラストラクチャーの整備を進めている。オークリッジ国立研究所(Oak Ridge National Laboratory; ORNL)のSummit、ローレンス・リバモア国立研究所(Lawrence Livermore National Laboratory; LLNL)のSierraはともにABCIとほぼ同時期、2018年に運用を開始している。

## (2) ABCI

AI橋渡しクラウド(以降、ABCIという)は、人工知能に関する「グローバル研究拠点整備事業」の一環として、東京大学柏IIキャンパス(千葉県柏市)の産総研柏サイトに新たに建設されたAIデータセンター棟[3]に導入された、AI・ビッグデータ処理向けの大規模かつ省電力な計算インフラストラクチャーである。ABCIの実効性能・省電力性能の高さは、HPLベンチマークの結果が示している。2018年6月のTOP500 List[4]において19.9 PFLOPSを記録し、世界5位かつ国内最高性能となった。また、電力最適化を行っていない段階ではあるものの12.054 GFLOPS/Wを記録し、Green500 List[5]の世界8位となった。2018年8月より本格運用を開始している。

ABCIは、我が国の人工知能技術開発のためのオープンなリーディングインフラストラクチャーとして、人工知能分野の最重要課題への挑戦。特に画像認識、音声認識、自然言語処理、種々の機械学習アルゴリズムやデータモデルの高度化、自動車/ロボットの自動運転/制御、創薬向け化合物推定、音声対話、自動翻訳等、幅広い分野での新たなアプリケーションの創出や、これらを支えるクラウド型計算システムの設計・運用ノウハウの民間への技術移転等、人工知能技術の社会実装の推進を目的としている。

### ① ABCIシステム

ABCIの外観を図2-8-4に示す。ABCIは、AIデータセンター棟が提供する電源、冷却システム、48Uラックを用いて導入された。

■ 図2-8-4 ABCIシステム外観

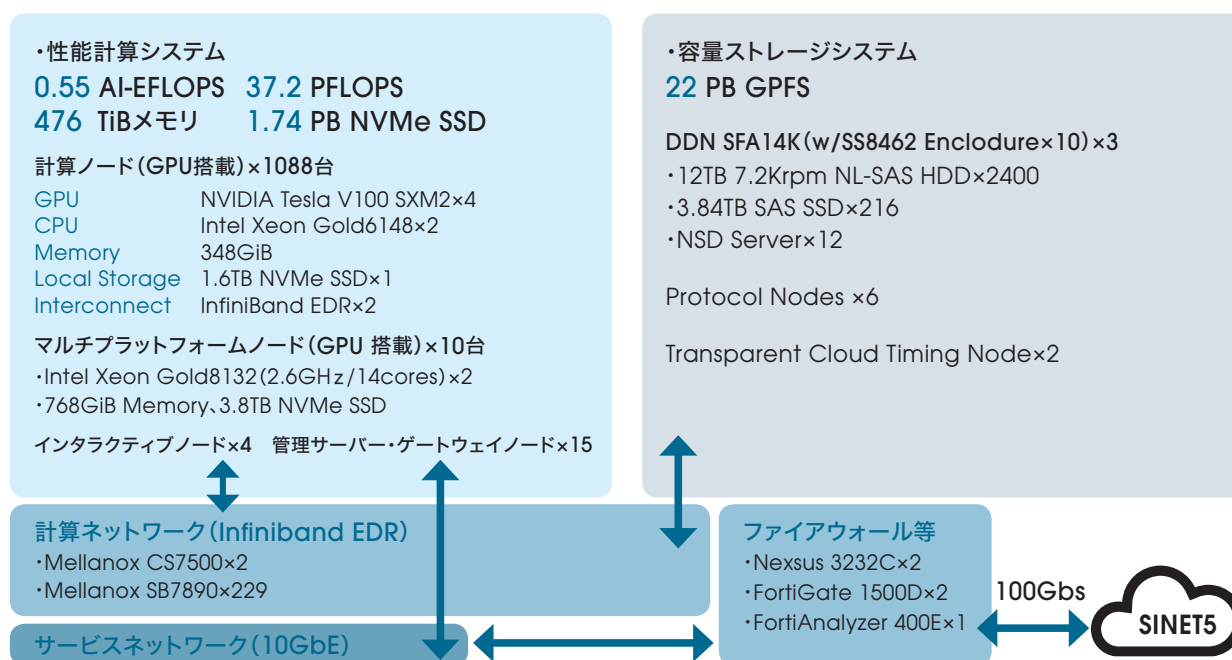


提供：産業技術総合研究所

ABCIシステムは、高性能計算システム、大容量ストレージシステム、計算ネットワーク、サービスネットワーク、管理ネットワーク、それらの補助的機器から構成されるハードウェアと、システムを最大限活用するためのソフトウェア群からなる。ハードウェアの概要を図2-8-5に示す。

計算ネットワークは、高性能計算システム及び大容量ストレージシステムを相互に接続する。サービスネットワークは、高性能計算システム及び大容量ストレージシステムの外部アクセスを必要とする機器群を接続するとともに、SINET5 100Gbpsに接続する。管理ネットワークは、各機器を管理・運用の用途のため接続する。

■ 図2-8-5 ABCIハードウェア構成



出典：産業技術総合研究所作成

以下、主要な構成要素である高性能計算システムと大容量ストレージシステムについて述べる。

## ② 高性能計算システム

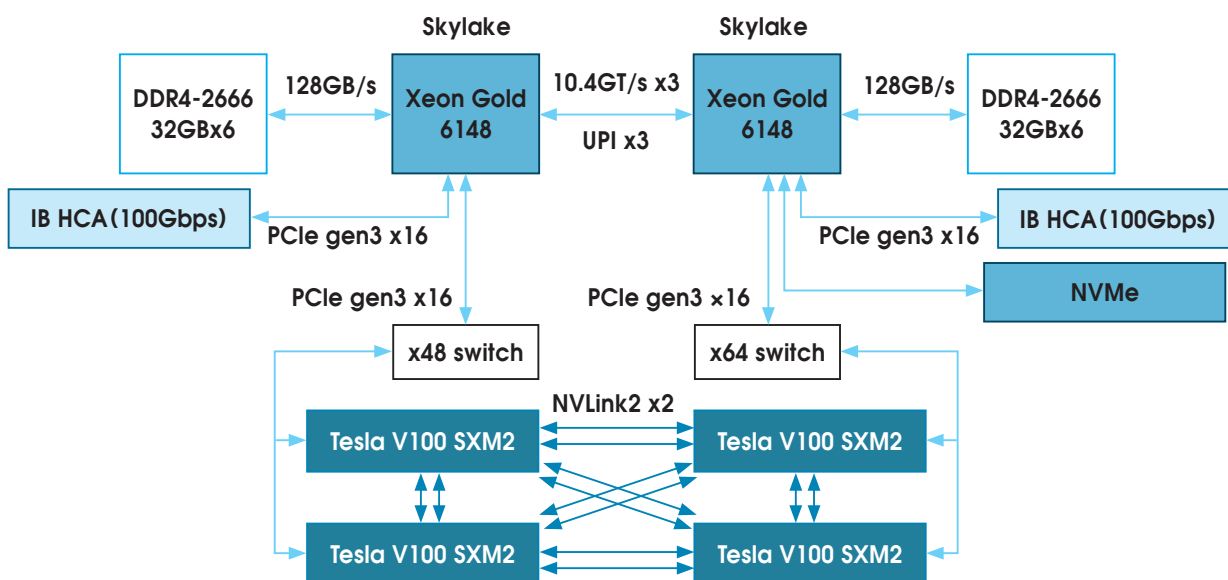
高性能計算システムは、計算ノード1,088台、マルチプラットフォームノード10台、インタラクティブノード4台、管理サーバー・ゲートウェイノード15台等からなる。以下では、計算ノードとそのインターコネクトを中心に説明する。

### 1) 計算ノード

ABCIの計算ノードは1,088台あり、そのすべての構成が同一である。FUJITSU Server PRIMERGY CX400 M4をベースとした2Uシャーシに、PRIMERGY CX2570をベースとした計算ノードを2台ずつ搭載した構成を基本とし、17シャーシ34ノードを48Uラックに搭載し、全体では32台のラックから構成される。

計算ノードの主要な構成は以下のとおり。ブロック図は図2-8-6に示す。

■ 図2-8-6 ABCI計算ノード構成



- CPU  
Intel Xeon Gold 6148 Processor (Skylake-EP、27.5MB Cache、2.40GHz、20 cores、1.536TF@FP64) × 2
- メインメモリ  
DDR4 2666MHz RDIMM (ECC) 384GiB (32GiB × 12) メモリバンド幅: 128GB/s × 2
- GPU  
— NVIDIA Tesla V100 SXM2 (5120 CUDA cores、16GiB HBM2、900GB/s、7.8TF@FP64、15.7TF@FP32、125TF@FP16) × 4
- ローカルSSD  
Intel SSD DC P4600 1.6TB u.2 × 1
- インターコネクト  
InfiniBand EDR (100Gbps) × 2

出典:産業技術総合研究所作成

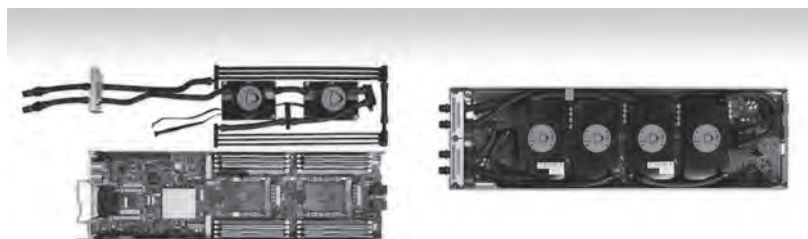
計算ノード全体では、NVIDIA Tesla V100 SXMが4,352基、Intel Xeon Gold 6148 Processorが2,176基、メモリ476TiB、メモリバンド幅4.19 PB/s、NVMe SSD 1.74PBとなる。また、理論ピーク性能では37.2 PFLOPS (FP64)、75PFLOPS (FP32)、0.55EFLOPS (HP)に相当する。

この計算ノード構成は、2017年8月に運用開始したTSUBAME3.0[6]の計算ノードを、約1年後の技術を用いて(すなわち、Skylake-EP、Voltaアーキテクチャーの恩恵を受けて)より安価で汎用性の高いPCサーバーをベースとした高密度パッケージングで再構成したものとも言える。

## 2) 計算ノードの冷却

参考文献[3]にあるようにAIデータセンター棟は、計算ラックまで32℃の冷却水を提供しており、CDUを介してラック内の計算ノードに分配される。計算ノードでは、高温になるCPU、GPU、メモリ等の基幹部品に取り付けられたコールドプレートを通じて冷却する。冷却しきれなかった熱はホットアイルに排出され、ラック上部に設置された(上記と同じ冷却水を利用する)ファンコイルユニットを用いて35℃程度まで冷却され、コールドアイルに排出される(図2-8-7)。

■図2-8-7 ABCI計算ノード



提供:産業技術総合研究所

## ③ その他システム

ディープラーニングを含む機械学習においては、大量のファイルI/Oが発生するため、しばしば学習処理スループットのボトルネックとなる。大容量ストレージシステムでもSAS SSD領域の一部を高速領域として利用しているが、容量とI/O性能の点で不十分である。このため、Burst Bufferのようなシステムの導入によりI/O性能をエンハンスすることは極めて重要である。

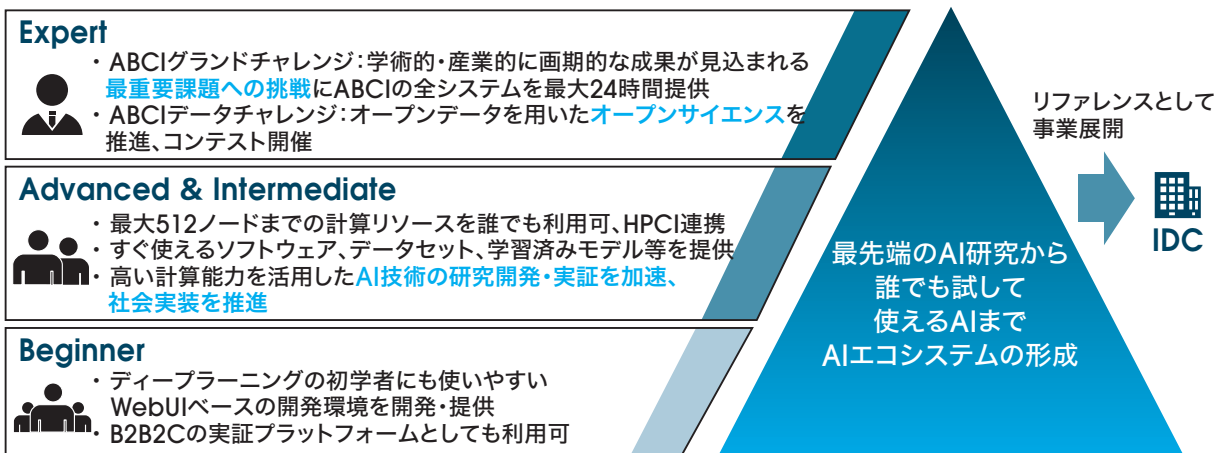
ABCIでは、BeeONDを用いて計算ノードのローカルSSDを用いたグローバルスクラッチ領域を構築できるようにした。また、今年度後半にはI/O性能に特化して、SSDのみで構成された「高性能キャンペーンストレージ」を導入し、現在の大容量ストレージシステムに代わるグローバルスクラッチ領域を構築する予定である。

また、人工知能技術の開発や応用に用いられるデータはしばしば非公開であり、機微な情報を含む場合が多い。2018年度後半をめどに、法令及び国際的なセキュリティ基準に沿ってデータを安全に管理することで産総研のみならずユーザー企業を含む他機関の保有する非公開データ等の保管を可能にする「セキュアオブジェクトストレージ」の導入も計画している。

## ④ ABCIのサービス設計・運用

ABCIは我が国の人工知能技術開発のためのオープンなリーディングインフラストラクチャーとして、人工知能分野の最重要課題への挑戦から、種々のアルゴリズム・データモデル開発、アプリケーション開発、はてはディープラーニングの初学者にいたるまで、幅広いレンジのユーザーとユースケースに対応し、人工知能技術の社会実装、実社会への橋渡しを推進する必要がある。

■ 図2-8-8 ABCIのユーザー・サービス階層



出典: 産業技術総合研究所作成

図2-8-8は、ユーザーのスキルレベル・使い方とその割合を想定したイメージである。Expertには、ABCIの全系を用いてトップノッチ成果を産出していくトップグループが位置する。産総研は、後述するABCIグランドチャレンジプログラムを主宰し、自らも国際的にも競争力の高い成果の蓄積を目指していく。

Advanced & Intermediateは、512ノードまでの中程度の計算リソースを利用した、画像認識、音声認識、自然言語処理、種々の機械学習アルゴリズムやデータモデルの高度化、自動車/ロボットの自動運転/制御、創薬向け化合物推定、音声対話、自動翻訳等、幅広い分野での新たなアプリケーションの創出に取り組むグループである。従来のHPCシステムのターゲットに近く、我々がサービスのベースラインと考えるグループでもある。

Beginnerは、NVIDIA DIGITSやSONY Neural Network Consoleなどoff-the-shelfの統合開発環境やトレーニングプログラムを通じてこれからディープラーニングに習熟していくグループで、ユーザー比率としては最大となる。また、ABCIを用いてエンドユーザー向けにB2Cサービスの提供を行う事業者に対しては、サービス実証のためのプログラムを用意することで、より使いやすい開発環境やアプリケーション実行環境の開発を促進することで、国内発のイノベーションの加速を支援している。

ABCIでは、先進的なシーズ開発や運用に向けた様々な取組みを進めている。以下では、推進中あるいは推進予定の運用に関わるプロジェクトについて述べる。

### ⑤ ABCIグランドチャレンジ

産総研は、莫大な演算能力によりはじめて可能になる人工知能分野の最重要課題への挑戦を支援するため、グランドチャレンジプログラムを実施している。本プログラムは、ABCIの全1,088ノード(4,352GPU)を最大24時間、1研究グループでの占有利用を可能にする公募型プログラムである。今年度は3回の実施を予定しており、いずれの回も2課題程度(第1回は3課題を採択)を採択する予定としている。採択課題については、チャレンジ実施前に小規模実行によるリハーサルを行う機会を提供する。また、利用料金はリハーサルを含めて無料としている。

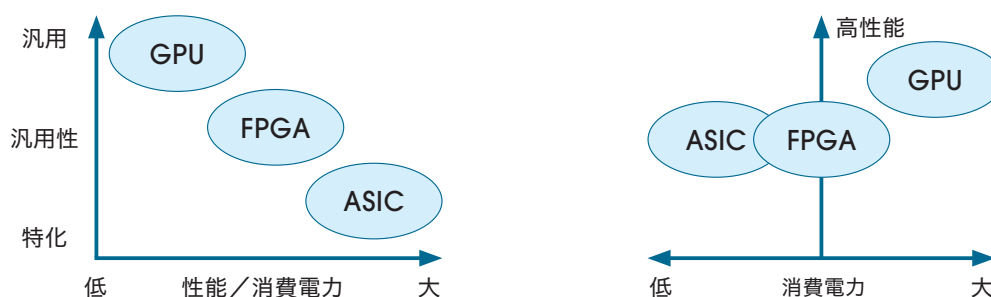
### 2.8.3 エッジ側基盤（推論用のプロセッサ技術と計算デバイスの動向）

ディープラーニングでは、インフラストラクチャーや計算デバイスに求められる性能と性能/電力比は、推論時と学習時で大きく異なる。特に、ネットワーク上のデータのトラフィックコストや、処理の時間的遅れ（レイテンシ）、使用可能なメモリ量、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。送付データ量が少なく、かつある程度の時間が許容できる場合、クラウド側の推論環境を利用したセミリアルタイムのアプリケーションが成立する。例えば、Googleや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・認識をエッジ側で行うことが求められる。エッジ側のプロセッサ・デバイスに対する要求として、電力制約に関しては、数十mW～数Wのレンジが要求される。もう一点のリアルタイム性に関しては、数msec～数十msecのレンジでレスポンスタイムが要求される。

#### （1）各計算デバイスの特徴

GPU、FPGA、専用チップ（ASIC）の違いを、性能対汎用性、性能対電力でそれぞれ可視化したものを図2-8-9に示す。互いの関係は、新技術の登場により容易に逆転する可能性がありうるので、あくまで以下に述べる説明の目安として見てほしい。

■ 図2-8-9 GPU、FPGA、専用チップ（ASIC）の特徴



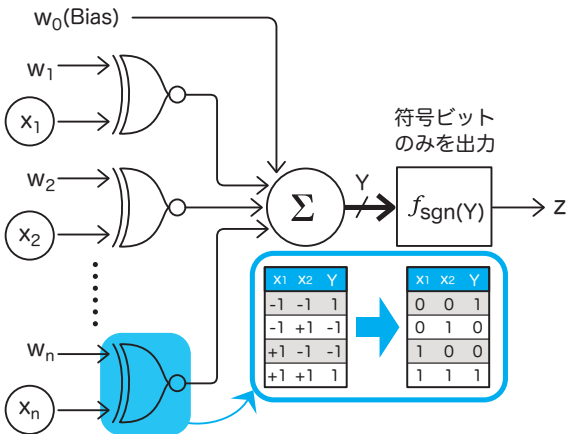
エッジで利用される推論用の計算デバイスには、学習用の計算デバイスと比較して、低消費電力であることが求められる場合が多い。従って、消費電力の大きいGPUではなく、搭載する計算ロジックを書き換えることができるFPGAや専用チップ（ASIC）のニーズが高まると考えられる。Intelは、FPGAを主力製品とするAlteraを買収し、CPUとFPGAを同一パッケージや同一ダイに統合した製品の開発を行っており、Xilinxは画像処理に特化した開発環境reVisionを発表している。これらはFPGAで推論に特化した回路を想定しており、学習までには対応していない。同様なツールはサードパーティからも発表・リリース予定があり、例えばTERADEEP（米国）、DeePhi Tech（中国）はすでにFPGAを採用した推論アクセラレーターをリリースしている。モデルを実装までつなげる開発環境や、LeapMind（日本）のようにFPGA向けのソリューション提供も盛んに行われている。



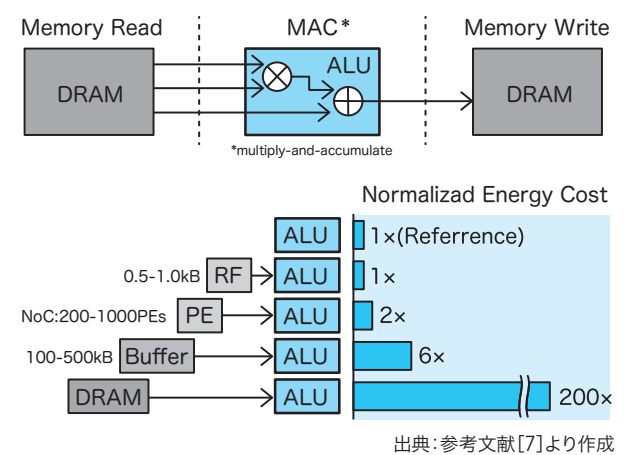
ディープラーニングの学習や推論において、必ずしも32ビットや64ビットの倍精度計算は必要なく、8ビット、4ビット、さらには1ビット(2値化) / 2ビット(3値化)での計算も可能との研究結果が相次いでおり、消費電力や計算性能の観点からFPGAを用いたプロトタイプの研究開発が進んでいる。

図2-8-10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。従って、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すべてのバッファを1チップに格納できるので、DRAM(Dynamic Random Access Memory)のような追加電力コストが生じる、オフチップは不要になる(図2-8-11)。

■ 図2-8-10 2値化ニューラルネットワークの概要



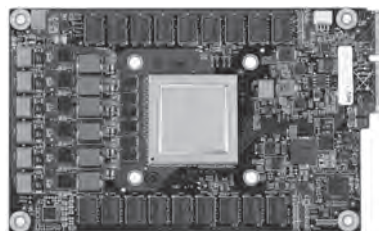
■ 図2-8-11 演算器と各種メモリの電力コスト



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 [8]。

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のディープラーニング専用チップであるTPU(図2-8-12)では、「計算精度の最適化」と「行列演算に特化した設計」により、従来のデバイスに比べ電力性能が30 ~ 80倍(TPU v1)であるとしている。また、

■ 図2-8-12 TPUボード



出典: Google Cloud Platform Japan Blogより

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きているとしている。また異なる開発フレームワーク間で、学習済みモデルを相互変換できるONNXのような標準化が進んだことにより、学習・推論の用途に応じて適切な開発フレームワークとデバイスの組み合わせを選択するといった流れが加速している。

以上をまとめると、組込みシステム・IoT向けの推論が汎用性の高いGPU型(SoCを含む)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ついていない。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のははずである。

しかしながら、現状では、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従できず早期に陳腐化してしまう可能性が高いため、FPGAやGPUの出番が続くと予想される。これらを巧みに組み合わせたSoCという方向性もあり、特に自動運転ではNVIDIA DRIVE Pegasusプラットフォームが注目されている。ここ数年は、主に上記3つの組込み型プロセッサが切磋琢磨する状況が続くと予想される。

## (2) 組込み型プロセッサ：推論の高速化・省エネルギー化

機械学習における推論に適したエッジ側の計算資源としては、5～15Wといった組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(ASIC)、GPU等の機械学習アクセラレーターを搭載したSoCや、DNNで必要とされる積和演算と相性のよいFPGAによる実装の研究開発が進められている。

### ●SoCによる実現

DNN用のGPUとCPU及び周辺回路を1チップ化(SoC)することで低消費電力、高性能な推論機能を実現する。本稿執筆時点でエッジ側でディープラーニングによる予測を行おうと考えた場合に、有力な候補になるのは、NVIDIA社のシステムオンチップ(SoC)のTegraである。Tegraはモバイル/IoT向けのプロセッサシリーズであり、ARMアーキテクチャーのCPUにNVIDIAのGPGPUを統合したSoCである。Tegraの利点は、ディープラーニングのデファクトスタンダードとなっているCUDAを用いたプログラムをそのまま実行できる点にある。最新のTegra X2(コード名Parker)は、半精度で1.5TFLOPS程度である。TDPは15Wである。SoCの応用例としては、同社の自動運転用プラットフォームである「NVIDIA DRIVE Pegasus」では、最新のVoltaコアを搭載したSoCであるTegra(Xavier)チップを2基組み込んで、複数のカメラやセンサーからの入力を、最大320TOPSの性能で処理できるとしている。

### ●FPGAによる実現[9]

FPGAは、プログラミング可能なハードウェアデバイスである。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。FPGAは後述するASICに比べると配線資源に柔軟性を持たせているため本質的に低速かつ高消費電力であり、同じアルゴリズムがASICで実現可能であれば利用する意味はない。しかし、現在のニューラルネットワークのように、アルゴリズムの研究が日進月歩で進展している場合、即座に実装して展開が可

能なFPGA実装に一定の価値がある。また、アプリケーションに完全に特化した回路構成を採用できるのもGPUに対するFPGAの利点であると考えられる。

FPGAで特に注目すべき手法として、低ビット幅精度の演算が挙げられる。ディープラーニングにおいては、高ビット幅を用いた浮動小数点演算は不要であることは古くから知られており、上述のArm MLのように離散化した整数値での演算が広く用いられつつあるが、一般にASICとして実装されているのは4～8ビットが多い。これに対し、研究レベルでは1ビットもしくは2ビット(2値化もしくは3値化)での学習・推論が可能なが示されている。しかし、これらの手法は、学習データセットやアプリケーションに対する汎用性が不明なため、現時点では、ASIC作成は陳腐化リスクが高い。これに対し、FPGAであれば有効であるアプリケーションに対してのみ利用することを前提に専用回路を設計できる。2値化や3値化に丸めると、乗算を小規模な論理回路で実現でき、回路面積の大幅な圧縮が可能だけでなく、重み行列のデータ量が小さくなりオフチップメモリへのアクセスが不要になる。従って、この技術はデータ依存性が強いディープラーニングでは、FPGAとの相性がよいと考えられ、活発に研究開発が進められている。日本でも、FPGA向けの開発ツールやキット販売を行っているLeapMindが米国調査サービスCB Insightsの「AI100」社に選出されるなど、活発な活動を行っている。

#### ●専用チップ(ASIC)による実現

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の電力効率性は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込み専用のASICを提案している。GoogleのTPUや、Intelが買収したMovidius(米国)の画像処理専用チップ等がある。

#### ●機械学習用のプロセッサ IPによる実現

組み込み用のプロセッサ IPに機械学習推論に適したDSP(Digital Signal Processor)を付与する動きは従前から各方面でなされているが、Arm(英国)のArm Machine Learning Processor(以下 Arm ML)もその一つである。Arm MLは、Arm社が提供を予定している機械学習向けのプロセッサで、メインとなるCPU、GPU、DSP等に対するアクセラレーターとして機能する。内部にSRAMを持ち、計算の大半をアクセラレーター内部で行うことで、計算の効率化と消費電力の低減を図っている。

Arm MLは、複数のCompute Engine(以下CE)と呼ばれるユニットで構成される。CEは、内部のSRAMを共有するとともに専用の内部ネットワークで結合される。CEは、MAC(Multiply-Accumulate) Engineと、PLE(Programmable Layer Engine)とで構成される。

MAC Engineは高速な積和演算を行い、PLEはアクティベーションの計算などのその他の操作を行う。汎用性の高いPLEを搭載することで、単純なCNNだけでなく、今後開発されるだろう新たなネットワークモデルへの対応を可能としている。

Arm MLの性能は16CEで4.6TOPs、7nmプロセスで実装した場合の電力比性能は、3TOPs/W以上とされている。上述のTegra X2は、単純に計算すると0.1TFLOPS/Wであり、これと比較すると非常に高いといえる(ただし上述のとおり、行う演算が異なる)。現在公開されている製品は、比較的小規模な監視カメラやスマートフォンなどのデバイスをターゲットにしたものだが、より大規模な製品への展開を予定している。

## 2.8.4 次世代AIインフラストラクチャー・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けたアーキテクチャーを備えたプロセッサ、あるいは脳型コンピューターと呼ばれるアーキテクチャーを構想し、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にもと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ーや、量子コンピューターについて述べる。

### (1) AIプロセッサの分類

従来型のコンピューターに対するAIコンピューター、脳型コンピューターの位置づけをノイマン型、非ノイマン型コンピューターの区別とともに図2-8-13に示す。AIの情報処理に向けたコンピューターをAIコンピューター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けたコンピューターも含まれる。LispやPrologに代表される論理プログラミング処理に特化した計算機もノイマン型のAIコンピューターと位置づけられる。

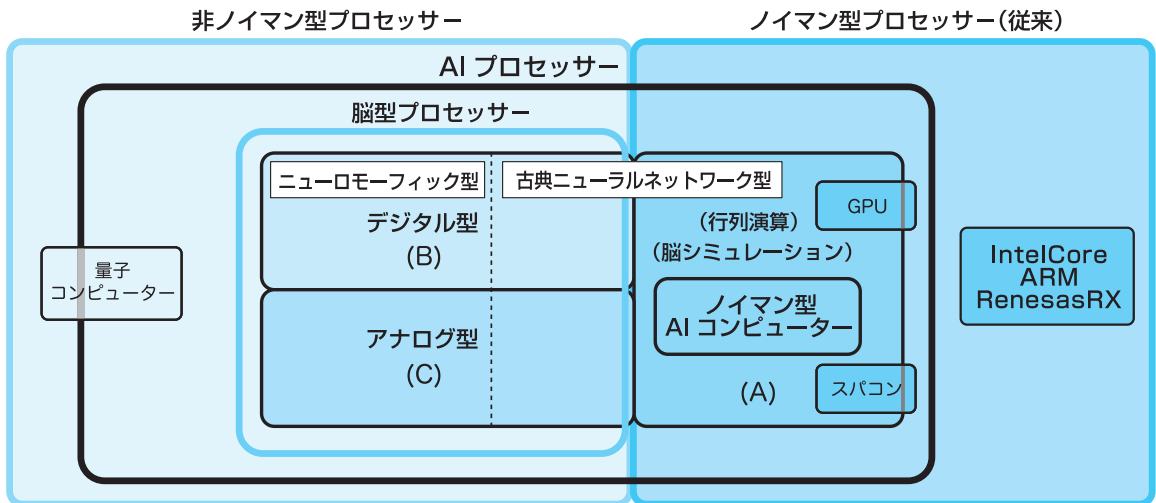
現在注目を浴びているAIコンピューターは、ニューラルネットワーク型、あるいはニューロモフィック型である。現代のコンピューターは、計算論的に万能マシンであるので、ディープラーニング計算をPCやスーパーコンピューター、あるいはGPGPUで実行させることが可能である。特にGPGPUは、数千を超えるコアを備えて高い並列性を発揮できるので、ノイマン型の中でも非ノイマン型に近いアーキテクチャーを持つ。

脳型プロセッサでは、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ーと同様にDRAM等のメモリを用いるデジタル型と、メモリストあるいはReRAM<sup>\*79</sup>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2-8-14に示す。

---

\*79 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。

■ 図2-8-13 AIコンピューター、脳型コンピューターの位置づけ



【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)】

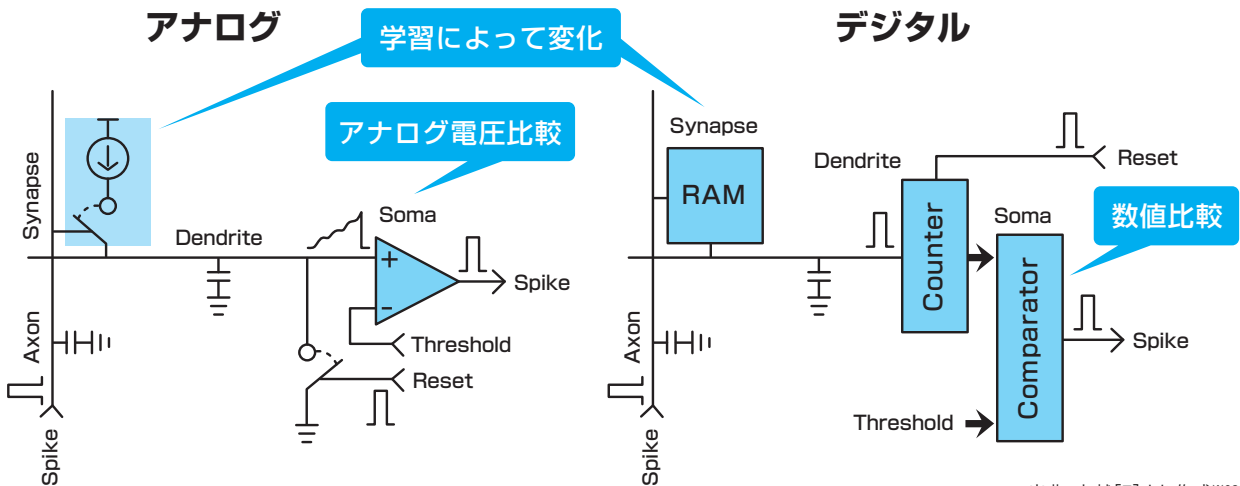
脳型プロセッサ：AI プロセッサの中で、ニューロン・シナプスのような脳機能を使った演算処理を行う

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)】

出典：NEDO技術戦略センター作成

■ 図2-8-14 シナプスとニューロン細胞体の実現回路例



出典：文献[7]より作成※80

※80 <[http://www.riken.jp/pr/topics/2013/20130802\\_2/](http://www.riken.jp/pr/topics/2013/20130802_2/)>

## (2) デジタル型のニューラルネットワークプロセッサ

本項では、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について述べる(表2-8-3)。データフロー型のアーキテクチャーを採用するものが多い。

■表2-8-3 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)

開発者	国	名称	アーキテクチャ	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70mW、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	2018年中出荷予定
Intel	米国	Loihi	データフロー型ニューロン13万個1.3億シナプス	・自己学習機能 ・プロセスは14nm FinFET	2017年9月発表
トプシステムズ	日本	SMYLEdeep	データフロー型 8コア 75MHz	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

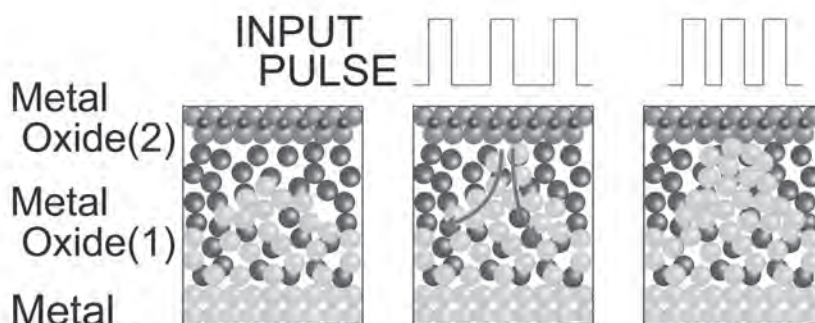
出典:各種公開資料より作成

表2-8-3の中では、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にすぎない。

## (3) 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。2008年にHPが「メモリストア」(memristor)を発見し、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。

■図2-8-15 電気抵抗メモリの変化



出典:「Aidevice」semiconportal Webサイト<sup>※81</sup>より

※81 「Aidevice」semiconportal Webサイト<[http://www.semiconportal.com/aist\\_aidevice/](http://www.semiconportal.com/aist_aidevice/)>

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2-8-15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2-8-4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。

■表2-8-4 アナログ型ニューラルネットワークプロセッサ

開発者	国	発表時期	可塑性	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し1チップ上に400万セルを集積化
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	メモリスター (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	メモリスター (Al2O3/TiO2)	CMOSニューロンの上に12×12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	メモリスター	32nmプロセスを用いた場合チップ面積85.4mm <sup>2</sup> メモリスター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	スピントロニクス素子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

出典: 各種公開資料より作成

#### (4) 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、現代のコンピューターで実行することが可能である。これに対し、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ーからの脱却と、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ーとは、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ーであり、現在主流となっている。このアーキテクチャーでは、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。

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ーは、まさにそのような構造になっている。

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、小脳などすべ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。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おりと同時に、消費電力の増大が問題となる。例えば、人間の脳は、消費エネルギー 20Wであるが、AlphaGo(碁)の消費電力は25万kWであるという。

すでに、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ーハードウェアは不適當であり、より生物の脳に近い脳型コンピューター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、推論時の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。誤差逆伝播学習は、計算量としては、推論時計算の数倍程度であるが、最適なニューロンの結合係数を求めるための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ーに任せ、エッジ側のプロセッサでは順方向計算に割り切るといった分担ができています。しかし、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と思われる。



## (5) 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

脳型コンピューターは、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。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、ニューラルネットワークのさらなる多層化や、ニューロン数を増やすだけで言語、計画、論理、創造など広範な人間的知能を獲得できると思われない。

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<sup>※82</sup>、オートエンコーダー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。

## (6) AIコンピューター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ーを活かすことで、従来のロジックCPUに比べて集積度を上げやすいこと、ReRAMやメモリスター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。さらに、ノイマン型アーキテクチャー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ーとすることが可能である。これらの考えにもとづき、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。例えば、米国防総省国防高等研究事業局DARPAは、「Hierarchical Identify Verify Exploit ; HIVE」と呼ばれるグラフ形状のデータ／知識ベースの処理のHWによる高速化の研究に取り組んでいる Webに代表される膨大な知識情報と連携した人工知能の実現には高速化は必須になるであろう。単純なニューロンモデルでなく、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の発展も考えられる。

---

※82 力学系的な定式化にもとづくニューラルネットワークの一種。EchoState NetworkやLiquid-state Machineなどの種類がある。

## (7) 量子計算機

さらに長期的には、量子計算の動向にも注目しておく必要がある。量子計算とは、量子論的な物質のふるまいを利用して高速な計算を実現しようとする試みである。深層学習の高速化に関連する。

量子計算の方式には主に3つに分類できる(表2-8-5)。

■表2-8-5 代表的な量子計算方式

	量子ゲート方式	量子アニーリング方式	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 of 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 量子ビット数 結線数	IBM / Google 50ビット/72ビット —/—	D-WAVE 2,000ビット 3,300	Nil-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(10mK) 超伝導量子回路	極低温(10mK) 超伝導量子回路	室温(300K) 光パラメトリック発振器ネットワーク	
応用範囲	因数分解(暗号解読)、量子シミュレーション	組み合わせ最適化、サンプリング(人工知能)	組み合わせ最適化、サンプリング(人工知能)	

出典:各種資料より作成

量子計算のハードウェア及び利用環境の開発も活発に行われ、実用化も始まっている。例えば量子アニーリング方式の代表的な例として、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ー「D-Wave One」(128量子ビット)を発表した、2017年には2,000量子ビットまで性能向上を行った(「D-Wave 2000Q」)。NASAや国防総省でプロジェクトが開始されており、民間での利用も始まっている<sup>\*83</sup>。日本でも東北大学と東工大協同の研究拠点に2019年秋に導入予定である。

AIと量子計算の関係としては、教師なし学習において有名な一手法である「ボルツマンマシン」に、量子コンピューターを利用するという提案(Quantum Deep Learning)がある。また、量子アニーリングは、最適化問題を高速に解くことができ、この特徴を利用して、深層学習の確率勾配法の中の最適化処理を量子計算で実現するなどが報告されている[10]。

### ◆参考文献

- [ 1 ] Fast Algorithms for Convolutional Neural Networks, < <https://arxiv.org/abs/1509.09308> >
- [ 2 ] Shmuel Winograd. "Arithmetic complexity of computations," volume 33. Siam, 1980.
- [ 3 ] 「AI橋渡しクラウド—AI Bridging Cloud Infrastructure (ABCI)—の構想」情報処理学会研究会, 2017-HPC-160
- [ 4 ] TOP-500 List < <https://www.top500.org/lists/top500/> >
- [ 5 ] Green500-List < <https://www.top500.org/green500/> >
- [ 6 ] <<http://www.gsic.fitech.ac.jp/tsubame3>>
- [ 7 ] Massachusetts Institute of Technology, Energy-Efficient Multimedia Systems Group, The Eyeriss Project, "Tutorial on Hardware Architectures for Deep Neural Networks"
- [ 8 ] R. Zhao et al., "Accelerating Binarized Convolutional Neural Networks with Software Programmable FPGAs," ISFPGA2017.
- [ 9 ] 中原啓貴「FPGAを用いたエッジ向けディープラーニングの研究開発動向」人工知能学会誌 Vol. 33 No. 1, pp.31—38, 2018.
- [ 10 ] 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.

※83 「VW、量子コンピューターで大都市の交通量を最適化」日経新聞2017/3/16

## 2.9 ▷ 標準化・オープンプラットフォーム

### 2.9.1 標準化

標準化とは、「もの」(製品)や「事柄」(サービス)に関して、社会生活上必要とされる「品質・性能の確保」、「安全性の確保」、「互換性の確保」、「試験・評価方法の統一」等を目的に、一定の基準を定める活動である。近年は、広大な市場を獲得するための企業そして国レベルの新たな戦略的ツールとして国際標準化が積極的に活用されてきている。

AIについても例外ではなく、AI自体に関わる標準化活動の他にも、社会や産業への影響度の大きさから、様々な応用分野でのAIの活用としての標準化が並行して進んでいる状況である。本項では、国際標準化機構 (ISO)、国際電気標準会議 (IEC) 及び米国電気電子技術者協会 (IEEE) の標準化活動を中心に紹介する。

#### (1) ISO/IEC JTC 1/SC 42 Artificial Intelligence

ISOとIECは情報関連の技術委員会の範囲が重なるため、1987年、合同の技術委員会であるISO/IEC JTC 1を発足し、それぞれの技術委員会を副委員会(SC)に移行した。欠番を含め、2018年9月現在、SC 42まで設置されている<sup>※84</sup>。各SCはユーザーインターフェース、セキュリティ技術、バイオメトリクスなど情報分野の要素技術を対象としており、重複を避けるために、あるSCでセキュリティ技術の標準化が必要となった場合にはセキュリティ技術のSCで扱うといった調整も行われる。2017年10月にロシア連邦のウラジオストクで開かれたISO/IEC JTC 1総会において、SC 42 Artificial Intelligenceの設置が決定された<sup>※85</sup>。2018年4月には北京において第一回総会が開催され、以下のWG(Working Group)及びSG(Study Group)の設置が決定されている。

- ・WG 1 Foundational Standards : 「AIのコンセプトと用語」及び「機械学習を用いたAIシステムのフレームワーク」の規格開発を行う。
- ・SG 1 on Computational approaches and characteristics of artificial intelligence systems : 異なる技術や既存の特殊なAIシステムなどを調べることによりAIシステムのアプローチや特徴を明らかにする。
- ・SG 2 on Trustworthiness : AIシステムの信頼性(透明性、検証性、説明性など)、工学的な問題や脅威・リスクの評価、AIシステムの堅牢性、精度、安全性、セキュリティ、プライバシーなどに向けたアプローチの調査などを行う。
- ・SG 3 on Use cases and applications : AI適用領域の同定や代表的なユースケースの収集などを行う。

また、SC 42の設置の際にJTC 1総会ならびに上位の委員会において十分な賛成が得られなかったJTC 1/WG 9 Big DataのSC 42への移行及び、“societal concerns”(アルゴリズムによるバイアスなど)の取扱いについて再検討・再提案されており、両方とも上位の委員会において承認さ

※84 “ISO/IEC JTC 1 Information technology”<<https://www.iso.org/committee/45020.html>>

※85 人工知能に関わる国際標準化がスタート(一般社団法人 情報処理学会)<[https://www.ipsj.or.jp/release/20180110\\_itscjnews.html](https://www.ipsj.or.jp/release/20180110_itscjnews.html)>

れている。

なお、2018年10月に米国で開催された第二回総会において、JIC 1/WG 9、SG 2、SG3は、それぞれWG 2、WG 3、WG 4となること、及びSG 1が継続となることが決議された。

## (2) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

IEEEは、AIや自律システムの倫理的配慮に関する国際イニシアティブ(IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems)を設置し、すべての技術者がAIや自律システムの設計や開発において倫理的配慮を行うことの優先度を高めるための活動を行っている。「倫理的に調和した設計(Ethically Aligned Design)」として知的な機械システムに対する恐怖や過度な期待を払拭すること、倫理的に調和や配慮された技術を作ることによって、イノベーションを促進することを目的とした、指針を発表している<sup>※86</sup>。

具体的な標準化活動としては、IEEE P7000(システム設計における倫理的懸念に対処するモデルプロセス)からP7013(自動顔分析技術の包括的ガイドライン)までの14のワーキンググループ(2018年8月時点)を設立し、設計や開発の倫理的配慮に関わる標準化の議論を進めている<sup>※87</sup>。またOpen Community for Ethics in Autonomous and Intelligent Systems(OCEANIS)というSDO(Standards Developing Organizations)間で作るフォーラムの設立を主導し、AIや自律システムを始めとするイノベーションにおける標準化の役割や、技術、倫理や価値に関する問題についてSDO間での対話を開始した<sup>※88</sup>。

## (3) その他の標準化活動

以上のように、国際標準化団体では、AIの枠組みの標準化や応用領域の中での標準化が進んできているが、昨今のディープラーニングの急速な発展に伴う、学習済みモデルの相互運用性などの技術面からの標準化活動も別途進んでいる。

ディープラーニングにおいては、GoogleのTensorflowやPreferred Networks(日本)のChainerに代表される主要IT企業が開発したオープンなフレームワークにもとづく開発が広く行われており、これらのフレームワーク間での開発された学習モデルすなわちニューラルネットワークのデータとしての相互運用性を確保する動きが広がっている。代表的な標準化活動としては、並列計算、グラフィックス、メディア用APIのオープンな標準規格を作成することを目的とする技術コンソーシアムであるクロノスグループ(Kronous Group)により策定された交換フォーマットNeural Network Exchange Format(NNEF)、Facebookなどが中心となって開発しオープンな仕様として公開されているネットワーク交換フォーマットOpen Neural Network Exchange Format(ONNX)がある。

NNEFの目的は、ディープラーニングの開発用のフレームワークで開発したモデルを、推論用チップ等のハードウェアベンダーのエンジンに橋渡しをすることであるとされており、一方、ONNXはもともとはPyTorchとCaffeの2つの研究コミュニティの間の学習モデルの相互運用性を確保する活動からスタートしたものである。NNEFは2017年12月にversion1.0がリリースされ、規格と同時に、CaffeやTensorflowで開発されたモデルからのコンバーターが公開された。一方ONNX

※86 Ethically Aligned Design <<https://ethicsinaction.ieee.org/>>

※87 「倫理的に調和した場の設計：責任ある研究・イノベーション実践例として」江間有紗、人工知能32巻5号(2017年9月)

※88 OCEANIS<<https://ethicsstandards.org/about/>>

もAmazonやMicrosoftの参加など広がりを見せている。

今後、あらゆる分野でAIの実用化や普及が進むことが期待される中、AIを利用するために必要となる標準の策定の重要性が増す。学習のためのインターフェース、学習精度の評価・保証制度、データ流通・利用のためのAPI、データ生成から利用における個人情報やセキュリティなどのそれぞれで並行して標準化が進むと想定される。内閣府の「人工知能技術戦略会議」がとりまとめた「人工知能技術戦略」<sup>※89</sup>でも示されているとおり、AIに関わる国際的な標準化の議論に我が国企業等も積極的に関与し、競争優位な市場環境を形成・拡大することが重要である。

## 2.9.2 オープンソース

以上のように国際標準化を巡り様々な活動がなされる一方、機械学習やディープラーニングはアルゴリズム、学習済みモデル、学習用データ、ソフトウェアコードなど開発した技術やデータ等を公開・共有することで、多様なプレイヤーを巻き込む連鎖的な技術開発が加速度的に進むとされている。それにもとづき、大手AI企業においては、機械学習やディープラーニングで用いられるフレームワークやライブラリをオープンソースソフトウェア(OSS)として公開し、ユーザーの集合知を集めたり、データやモデルやノウハウを参加者が自由に交換し互いに協力し合うエコシステムやプラットフォーム<sup>※90</sup>を構築する動きがある(表2-9-1)。

プラットフォームを利用することで、AIの最新研究成果を誰でも簡単に利用できる環境が整備されつつあり、いわゆる「AIの民主化<sup>※91</sup>」の一翼を企業活動の一環として担っているともいえる。

■表2-9-1 機械学習やディープラーニングに関わる主要なOSS

フレームワーク名	Caffe	TensorFlow	Chainer	CNTK	MXNet
言語	C++/Python/ Matlab	C++/Python	Python	C++	Python他
開発元	Berkeley	Google	PFN	Microsoft	Apatch財団
特徴	・画像認識向き、 高速	・ユーザー数最多	・使いやすい	・RNNに強い	・AWSと相性良い
	・コーディングは 面倒	・DNNの構造が 分かっている人 向き	・動的 ネットワーク構築 (define by run)	・豊富なモデル ライブラリ	・高速で柔軟性の 高いライブラリ

出典:各種公開資料より作成

Caffeは、カリフォルニア大学バークレー校の研究センターであるBVLICが中心となって開発しているPython言語向けの代表的なディープラーニングライブラリである。C++で実装され、GPUに対応しているため、高速な計算処理が可能である。“Caffe is a community”というキャッチコピーもあるようにコミュニティにより活発に開発が行われている。

GoogleのTensorFlowは同社のMachine Intelligence研究所が機械学習やディープラーニングの研究を行う目的で開発されたものであり、その成果がApache 2.0オープンソースライセンスの下で、学生、研究者、エンジニア、開発者等に向けて広く公開されている。複数のGPUへの対応

※89 「人工知能技術戦略(平成29年3月31日)」<<http://www.nedo.go.jp/content/100862413.pdf>>

※90 「個人や企業などのプレイヤーが参加することではじめて価値を持ち、また参加者が増えれば増えるほど価値が増幅する、主にIT企業が展開するインターネットサービス」『ザ・プラットフォーム』尾原和啓、NHK出版、2015。

※91 専門家がいらない企業でも、個人としてでも、「誰もがAI技術を使えるようになること」を「AIの民主化」と呼ぶ。GoogleのAI部門の開発責任者であったFei-Fei Li氏の発言が源流とされている。

やPCやモバイル端末でも動作する点が特徴的であり、最も人気のあるフレームワークとなっている。

日本では、Preferred Networksが中心となって開発しているChainerフレームワークがあり、動的計算グラフの主要なニューラルネットワークフレームワークとして様々な長さの入力が可能なネットワークという特徴を持つため、自然言語処理作業に人気が高い。

MicrosoftのCNTK (Computational Network Toolkit) も社内用に開発されたライブラリを公開したもので、当初はC++言語のインターフェースのみを提供していたが、version 2.0からはAI分野でよく用いられるPython言語へのAPIを提供し、またモデルの相互運用性のための標準であるONNXにもversion 2.5から正式に対応するなど、積極的にオープンソースコミュニティへの参加をアピールしている。

AmazonはもともとDSSTNEというライブラリを公開していたが、ワシントン大学とカーネギーメロン大学で開発されたライブラリMXNetを自社クラウドの推奨フレームワークとして採用しオープンな開発を進めている。MXNetがスケーラブルであることと、学習済みモデルが軽量で比較的处理能力の低いデバイスでも利用できるという特徴を持っている。

他にも、Intelによる「BigDL」と称する分散型ディープラーニングライブラリのOSS公開、中国のBaiduが自社の検索ランキングやターゲット広告、画像検索、翻訳等に搭載されている分散型ディープラーニングのためのフレームワークPaddlePaddle (Parallel Distributed Deep Learning) を公開している。

このように、大手IT企業、スタートアップ企業等が、自社で開発したフレームワークやライブラリをOSSとして外部に公開している。学習用データや学習済みモデルなども合わせて公開することで、外部の企業や研究者を巻き込んだエコシステム、集合知のプラットフォームの形成を狙っている。

### 2.9.3 クラウドを利用したAIと開発環境

ここでは、Google、Azure、Amazonに代表されるクラウドサービスを活用した、クラウドベース開発環境について紹介する。

企業等が機械学習やディープラーニングを利用するためには、学習のための膨大なデータを管理するストレージ、その膨大なデータを計算処理するためのサーバー資源等を確保することが必要となる。ディープラーニングでは膨大なデータを高速に処理するために、GPUを大量に搭載した計算機環境が必須であるが、これを単独の企業でそろえることは難しい。またGPUや学習用のソフトウェア技術の進展が速いので、企業の通常の設定導入サイクルに従っていると導入時にすでに旧式、最悪の場合は最新のソフトウェアが動作しないという事態も生じかねない。そこでクラウドの利用機会が増える。クラウドサービスを提供する各社も、AI開発向けのハードウェアや開発環境を整備してサービスを提供している。サービスにはIaaS型、SaaS型の2種類がある。

#### (1) IaaS型クラウドによるAI開発

IaaS型クラウドは仮想化したハードウェアそのものを提供する。クラウドをAI開発に活用する利点は、最新のアクセラレーター (GPU) を時間単位の課金で利用できることにある。また、多くのクラウドベンダーは、AI開発環境に適したOSイメージのひな型を提供しているため、これらを利用することで、ソフトウェアスタックのセットアップコストを低減することができる。

一般に、クラウドの利用価格は、保有する場合と比較すると電気代を考慮に入れても高額だが、複数台を同時に利用することでハイパーパラメーターチューニングのターンアラウンドタイムを短くす

ることができる、常に最新のアーキテクチャーを利用できる、初期投入コストが小さい、などのメリットがある。

表2-9-2に2018年7月現在利用可能な、AIに適したアクセラレーター (GPU) を持つIaaSサービスの例を示す。主要なクラウドベンダーの多くがAIに適したCUDAやNVIDIA社製の最新GPUを利用できるノードを提供している。事業者によっては高速なネットワークも併せて提供しており、効率的な並列機械学習が可能となっている。

■表2-9-2 代表的なAIに適したIaaSサービス

サービス名	タイプ名	GPU機種	GPU数
Amazon Web Services	P2	Tesla K80	½-8
	P3	Tesla V100	1-8
Google Cloud Platform	K80	Tesla K80	½-4
	P100	Tesla P100	1-4
Microsoft Azure	NCv2	Tesla P100	1-4
	NCv3	Tesla V100	1-4
さくらインターネット 高火力	Tesla P40	Tesla P40	1
	Tesla P100	Tesla P100	1
	Tesla V100	Tesla V100	1

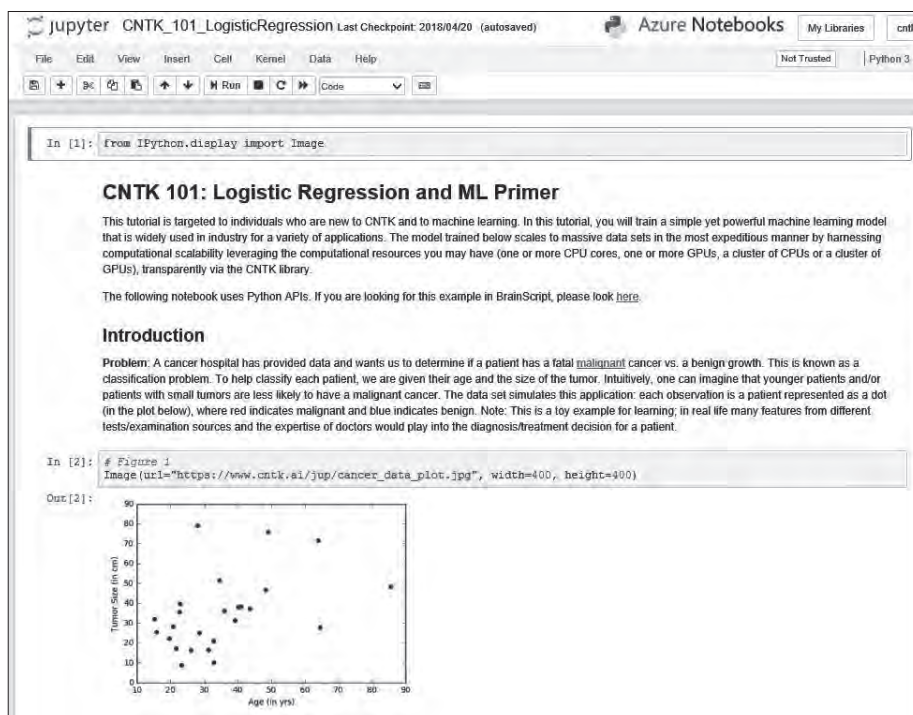
## (2) SaaS型クラウドによるAI開発

さらに、計算資源のほかに、機械学習やディープラーニング等のミドルウェア (機能) を組み合わせて活用するスキルが求められ、これは企業ですでに使っている企業内開発環境の整備に必要なスキルとは異なったものである。そこで、自前で計算機環境を整備して利用する形態から、機械学習等のAI機能を搭載したクラウドサービス (ここではクラウドAIと呼ぶ) を利用する形態への乗り換えが広がっている。

クラウドAIには、音声認識や画像認識、言語翻訳等のディープラーニング等を用いた機能のAPI (Application Program Interface) が搭載されており、利用企業は必要なAPIを通じて必要な機能を扱うことができる。従量課金制度への支払いルールが社内に整備できれば、従量課金の特性を活かして、PoC (Proof of Concept) などスモールスタート用途にも適しているともいえる。

さらにこれらのクラウドサービスでは、初学者がAIに関わるデータやPythonなどの基盤ソフトウェア・AIライブラリに簡単に触れられるように、限定された範囲 (ディスク容量や利用時間で制限) で利用できるインタラクティブなAI実行環境を無償で提供している。例えば、MicrosoftはAzure Notebooksとして、AI開発者に人気のあるPythonを手軽に利用できるJupyter Notebook環境を提供している。Jupyter NotebookはWebブラウザからページを編集するようにPythonをインタラクティブに入出力できる環境であり、計算過程および結果をWebページとして保存できるので好きなときに保存し好きなときに継続でき、できあがったWebページを技術ドキュメントとして活用することもできる (図2-9-1)。

■ 図2-9-1 notebookの例



出典: Microsoft Azure Notebooks

Azure NotebooksはこのWebページをMicrosoftアカウントを持つすべてのユーザーに開放し、当該Webページにアクセスするとクラウド側でAIライブラリが整備された専用の仮想サーバーが起動し、チュートリアルなどの例題をすぐに試すことができる。さらにオープンソースプロジェクトの最大の開発リポジトリであるGitHubとも連携し、相互利用が可能となっている。

Webブラウザとネットワーク接続さえあれば専用のPCを持たなくてもタブレットやスマートフォンから利用できるのも、誰でもAI技術を使えるようになるという意味で、「AIの民主化」に寄与している。また、クラウドベンダーからすると開発者のすそ野を広げることによる、エコシステムの拡大戦略でもあり、SaaS型クラウドの無料サービス提供を通じて自社APIへ誘導している。表2-9-3に代表的なクラウドサービスとAPIを整理する。



■表2-9-3 主要なAIクラウドサービス

組織名	サービス名	概要
Google	Google Cloud Machine Learning	機械学習用途のクラウドサービス。事前学習済みモデルも提供。Speech API (音声認識)、Vision API (画像分析)、Translate API (翻訳)、Natural Language (テキスト分析)、Jobs API (仕事検索機能) ・機械学習サービス: Machine Learning Engine ・対話型ツール: Google Colaboratory
Amazon	Amazon AI	AI利用のための以下の機能が提供されている。 ・Amazon Rekognition (画像認識)、Amazon Polly (文章から音声への変換) ・機械学習サービス: Amazon Machine Learning ・対話型ツール: Amazon SageMaker
Microsoft	Azure Machine Learning	視覚や言語、音声、認知、検索などを扱うコグニティブ関連サービス「Microsoft Cognitive Services」は29種類のサービスを用意。視覚の分野では、画像分析 (Computer Vision API)、顔認識 (Face API)、感情認識 (Emotion API) など8種を提供。言語・音声の分野では、意図解釈 (Language Understanding)、テキスト翻訳 (Translator Text API)、リアルタイムの音声翻訳 (Translator Speech API) などが日本語対応として提供 ・機械学習サービス: Azure Machine Learning ・対話型ツール: Azure Notebooks
IBM	Watson Data Platform	高速なデータ取り込みエンジン (100GB / 秒) や機械学習機能を提供するIBM Cloudベースのデータプラットフォーム。データ活用に関わる4つの専門職 (データエンジニア、データサイエンティスト、ビジネスアナリスト、アプリケーション開発者) のコラボレーションを可能にするデータ分析基盤を提供 ・機械学習サービス: IBM Watson Data Platform ・対話型ツール: Data Science Experience

出典:各種公開情報より作成

## 2.9.4 共有データセット・共有モデル

前項でも述べたように、機械学習やディープラーニングはアルゴリズム、学習済みモデル、学習用データ、ソフトウェアコードなど開発した技術やデータ等を公開・共有することで、多様なプレイヤーを巻き込む連鎖的な技術開発が加速度的に進むというエコシステム上の好循環が有効に働く。本項では、学習用データの公開・共有 (共有データセット) や学習済みモデルの公開・共有 (共有モデル) に関わる動向等について述べる。

共有データセット・共有モデルに関する公的なAI用クラウドに対する具体的な要件や機能については、『AI白書2017』(「2.3.5 共有データセット・共有モデル」)を参照されたい。

### (1) 共有データセット

#### ① 既存の共有データセット

現在、使用可能な共有データセットとしては、画像認識用データセットの「ImageNet」、手書き数字認識用の「MNIST」、画像をピクセル単位で意味づけし、領域を識別する画像セマンティックセグメンテーション用の「MS COCO」などがある。AIの性能を決定づけるのがデータセットであり、そのため世界中の様々な研究機関がデータセットを公開している。表2-9-4に主要な共有データセットを示す。

■表2-9-4 主要な共有データセット

データセット名	説明
ImageNet	スタンフォード大学がインターネット上から画像を集め分類したデータセット。一般画像認識用に用いられる。ImageNetを利用して画像検出、識別精度を競うThe ImageNet Large Scale Visual Recognition Challenge (e ILSVRC) などコンテストも開かれる
MNIST	ニューヨーク大学のYann LeCunらがまとめた手書き数字を集めたデータセット。機械学習の入門用データセットとしてよく用いられる
MS COCO	Microsoftが作成した画像セマンティックセグメンテーション用データセット
Pascal VOC	欧州のAI関連研究コミュニティのPattern Analysis Statistical Modeling and Computational Learning (PASCAL) がまとめた画像セマンティックセグメンテーション用データセット
WMT15	機械翻訳に関する国際会議のWorkshop on Statistical Machine Translation (WMT) がまとめた英語と仏語、独語、チェコ語、ロシア語、スペイン語、フィンランド語などの対訳集
Cornell Movie-Dialogs Corpus	コーネル大学 (Cornell University、米国) が600以上の映画の字幕をまとめたデータセット
VGG Face Dataset	オックスフォード大学Visual Geometry Groupがまとめた顔画像のデータセット
Human Pose Estimate Dataset	オックスフォード大学Visual Geometry Groupがまとめた人間のポーズ推定を学習するためのデータセット
Oxford Buildings Dataset	オックスフォード大学Visual Geometry Groupがまとめた建物画像のデータセット
BBC-Oxford Lip Reading Dataset	British Broadcasting Corporation (BBC、英国) とオックスフォード大学がまとめた唇から言葉を読み取るためのデータセット
Text Localisation Dataset	オックスフォード大学がまとめた、写真の中にある文字の位置を学習させるためのデータセット

出典：各種公開資料より作成

## ②既存の共有データセットの課題

現在広く頒布されているデータセットには、課題が2つある。一つは著作権や肖像権の課題である。現在はフェアユース（公正な利用）に近い形で運用されているが、これを企業が用いて学習させたものをデータセットとして、業務や製品に使うという問題への指摘がある。

GoogleやMicrosoft、Appleなどは、著作権などの知的財産権の処理をグレーゾーンとして捉え、自社のサービス用AIに共有データセットを学習させたうえで、有料のディープラーニングサービスの提供に踏み切っているが、コンプライアンス上問題があるという指摘がある。著作権などの知的財産権に関わる課題については、「4.2 知的財産」で説明する。

もう一つの課題は、データの地域依存性である。現在ある共有データセットはほとんど欧米で作られており、例えば交通標識や郵便ポストの色形など、日本の特徴に合わせたデータではないので、そのままでは日本では使えない。

## (2) 共有モデル

### ①既存の学習済みモデル(ニューラルネットワークモデル)

エコシステム上の好循環を生み出すもう一つの方法は、共有の学習済みモデル(共有モデル)を活用することである。学習済みモデルとは、すでに十分な時間、膨大な学習データセットを使用して学習(訓練)させたもので、米国ではカリフォルニア大学バークレー校の画像及び学習センター(Berkeley Vision and Learning Center ; BVLC)によるModel Zooで、多数の学習済みモデルが公開されている。Model Zooで公開されている主要な学習済みモデルを表2-9-5に示す。

■表2-9-5 BVLCのModel Zooで公開されている学習済みモデルの例

モデル名	説明
AlexNet	最初期の画像認識用ニューラルネットワークモデル
GoogLeNet	Googleが開発した画像認識用ニューラルネットワークをBVLCが独自に学習したもの
VGG-19	ILSVRC2012のデータにおいて7.5%のTop5エラー率を達成した画像認識用ニューラルネットワークモデル
Places CNN	マサチューセッツ工科大学が開発、学習した写真から場所を推定するニューラルネットワークモデル
FCNs	セマンティックセグメンテーションを行う完全畳み込みネットワーク
Age and Gender	年齢及び性別を認識するニューラルネットワークモデル
GoogleNet_cars	GoogLeNetを利用して自動車の種別を認識するニューラルネットワークモデル
SegNet	セマンティックセグメンテーション用のニューラルネットワークモデル
Holistically-Nested Edge Detection	輪郭検出用のニューラルネットワークモデル
Video2Text_VGG	動画からテキストを自動生成するニューラルネットワークモデル
VGG Face	VGGを使用した顔検出用のニューラルネットワークモデル
Emotion Recognition	顔画像から感情を認識するニューラルネットワークネットワークモデル
ResNets	Microsoftが開発した152層に及ぶニューラルネットワークモデル
Deep Hand	手の形を認識するニューラルネットワークモデル
DeepYeast	顕微鏡写真から体内細胞を認識するニューラルネットワークモデル

出典: 各種公開資料より作成

高性能なAIを作るためには、こうした公開されているモデルを元に、転移学習(ファインチューニングと呼ばれる)を独自のデータセットに対して適用するのが実用上効率的であることが知られている。

BVLCのModel Zooでは、BVLC自身が学習した学習済みモデルに加え、MITやMicrosoftなど、様々な研究機関が独自に学習させた学習済みモデルが多数公開されており、事実上のディープラーニング研究者たちのハブとなる役割を果たしてきた。「2.9.2 オープンソース」で紹介したフレームワークにおいても、各々の学習済みモデルが用意されている。

## ②学習済みモデル間の相互運用性の問題

モデル再利用の観点からは、フレームワーク間で学習済みモデルの相互運用性がないことが、従来から問題とされてきた。例えば、BVLC Model Zooで公開されている学習済みモデルは、BVLCが開発したディープラーニングツールである「Caffe」用のモデルに限られており、他のフレームワークからは直接再利用が困難な状況である。しかし各種フレームワークでも開発元あるいはOSSらしくユーザー有志により、モデルの移植が行われており、逆にフレームワークの得意とする領域のモデルの蓄積も個別に進んでいる。また前述のONNXやNNEFのようにモデル間の相互運用性を目指す活動も盛んであり、モデルの共有財産としての断片化はいずれ解消されると思われる。

## 2.9.5 オープンプラットフォーム、エコシステム

ここでは、クラウドファンディングのような環境でデータとモデルをコミュニティベースでそろえる、Kaggleについて紹介する。

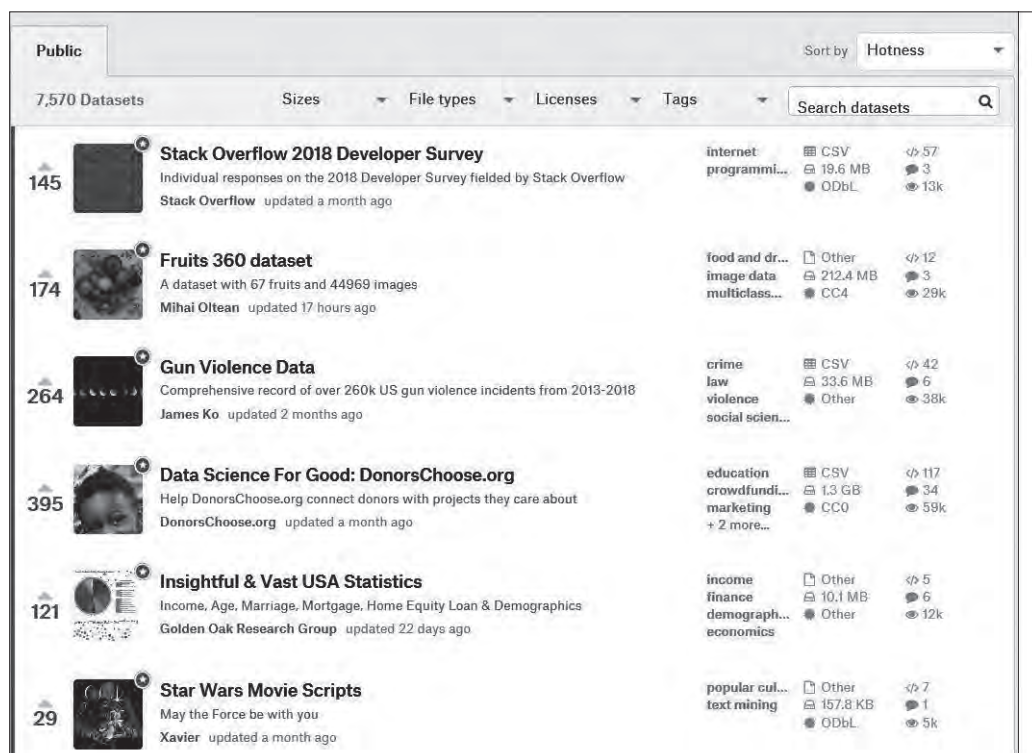
Kaggleは2010年にアメリカで設立された、世界中の企業や研究者がデータやモデルを投稿するプラットフォームである。世界中の機械学習・データサイエンスに携わっている約60万人が集まるコミュニティであり、企業や政府などの組織とデータ分析のプロであるデータサイエンティスト/機械学習エンジニアをつなげるプラットフォームとなっている。単純なマッチングではなく、「Competition (コンペ)」と呼ばれる、企業や政府が課題を提示し、参加者がオープンな場で競争を行い、賞金と引き換えに企業がもっともすぐれたモデルを買い取るという仕組みを提供しているのが、Kaggleの大きな特徴となっている。

初心者向けのチュートリアルや、掲示板なども整備され、またKernelと呼ばれる予測されたデータのファイルだけでなく、ファイルの事前処理を含むすべてのモデルのコードが説明付きで公開され、これらを活用し、参加者間での切磋琢磨を支援することで、集合知のプラットフォームになっている。これらの活動を通じ、実課題に即したデータセット、だれでもアクセスできるデータセットが整備されることになる<sup>※92</sup> (図2-9-2)。

---

※92 KaggleWebサイト<<https://www.kaggle.com/datasets>>

■ 図2-9-2 Kaggleのデータセットの例



出典: KaggleWebサイト

Kaggleへの参加実績は、世界的な企業への就職時の能力評価基準となっている。また逆に自社の能力の高さを宣伝するために、社員を就業時間中にKaggleに参加することを認めている会社もあり、Kaggleで腕を磨いた最先端のデータサイエンス人材は「Kaggler (カグラー)」と呼ばれて熱い視線が注がれている。Kagglerは実績に応じてランクづけ<sup>※93</sup>されており、一部の先端的な企業ではKaggleのランクに応じて、Kaggleの業務時間内利用許可比率を定めており、トップ入賞5回の最高ランクではすべての業務時間をKaggleのコンペに使うことが認められる<sup>※94</sup>。

※93 コンペ上位入賞者にメダルが授与されたり、メダル数に応じて、称号Grandmaster、master、Expert、Contributor、Novice等が与えられる。

※94 激化する「Kaggle人材」データサイエンティスト争奪戦<<https://www.businessinsider.jp/post-164895>>

## 2.10▷ 各国の研究開発

### 2.10.1 各国の研究開発の現状

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 (Artificial Intelligence ; 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、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。民間企業でも、自動運転や生産ロボットなど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、応用研究をさら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深層学習に関する取組みが早い段階から展開されている。2006年に深層学習の研究の発端となった論文を執筆したGeoffrey Hintonは、University of Toronto (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、深層学習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、アドバイザー等に関わっており、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、中国社会におけるAI社会実装が進んでいる。

米国政府は、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た。2016年5月、米国政府は、“Preparing for the future of artificial intelligence”というレポートを発表し、AIの研究開発の方向性を示した。2018年5月、米国政府は“White House Summit on Artificial Intelligence”を開催し、①AIにおける米国のリーダーシップ維持、②米国労働者を支援、③公的研究開発の推進、④イノベーションへの障壁を取り除くこと、を目標とすることと、AI委員会を組織することを示した。そのほか、BRAIN Initiative (National Institutes of Health ; NIH)、Precision Medicine Initiative (同) や国防高等研究計画局 (Defense Advanced Research Projects Agency ; DARPA) は、SyNAPSEプログラム等AIと関連の深い分野の研究開発を推進しており、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。

カナダは、2017年3月にいち早くAI戦略である“Pan-Canadian Artificial Intelligence Strategy”の中で、1億2,500万カナダドルを拠出することを公表した。同戦略では、優れた研究者の増員や研究能力向上を含むカナダのAIエコシステム構築を行うとした。

欧州委員会は、Framework Programme 7 (FP7)、その後のHorizon 2020の中で、AI技術関連プロジェクトを推進してきた。2018年4月、同委員会は“Communication on Artificial Intelligence”を採択し、研究投資額の増額にコミットするとともに、①EUの技術及び産業におけるAIの能力向上及び経済成長、②AIによる社会変革への準備、③AIに関する倫理及び制度設計の保証を目指す、との目標を示した。さらに、2018年末までに加盟国と調整して今後の計画を策定する予定である。

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AI研究が盛んであるとともに、先端的な深層学習の研究開発を行っているDeepMind<sup>\*95</sup>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 (Deutsches Forschungszentrum für Künstliche Intelligenz GmbH ; 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 (Max Planck Institutes) 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。深層学習への対応は米国に先を越されたと言わざるを得ないが、これまでHorizon2020<sup>\*96</sup>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。2017年7月には政府は「次世代人工知能発展計画」を、2017年12月に「次世代人工知能産業の発展促進に関する三年行動計画 (2018～2020年)」を相次いで発表し、2020年までにAI重点製品の大量生産、重要な基礎能力の全面的強化、スマート製造の発展深化、AI産業の支援体制の確立等を通じた重点分野の国際競争力の強化、AIと実体経済の融合深化等を目指すとの目標を達成するための4タスクが示された。また、Baidu、Alibaba、Tencent等の情報系企業はAIを活用したビジネスを積極的に進めている。

米国・中国・欧州の政策及びプロジェクトについては第4章にて詳細に紹介されるので、本節では特に我が国の政策・プロジェクトの現状、国防高等研究計画局 (DARPA) の研究プロジェクトや第3次AIブームの起爆剤でもある「ディープラーニング (深層学習)」の研究の起点である地の利を活かすカナダの政策・プロジェクトの概要について紹介し、さらに、各国の民間企業及び研究機関を中心としたAI技術の研究開発動向を紹介する。

## (1) 各国の政策・プロジェクトの現状

### 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」<sup>\*97</sup> (平成28年1月閣議決定) 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づけ、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、生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みにより、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し、その実現に向けた一連の取組みをさらに深化させつつ「Society 5.0」<sup>\*98</sup>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した。

第5回「未来投資に向けた官民対話」(平成28年4月)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、上記3省のそれぞれが所管するAI研究のため

---

※95 2014年にGoogleに買収された。

※96 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。

※97 内閣府Webサイト「科学技術基本計画」<<http://www8.cao.go.jp/cstp/kihonkeikaku/index5.html>>

※98 狩猟社会、農耕社会、工業社会、情報社会に続くような新たな社会を生み出す変革を科学技術イノベーションが先導していく、という意味が込められている。

のセンター(情報通信研究機構：NICT<総務省>、理化学研究所革新知能統合センター：API<文部科学省>、産業技術総合研究所人工知能研究センター：AIRC<経済産業省>)の研究の総合調整を行う場として研究連携会議を設置するとともに、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を設置して議論を行っている<sup>\*99</sup>。平成29年3月には「人工知能技術戦略」<sup>\*100</sup>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」<sup>\*101</sup>を策定した。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行い(表2-10-1)、さらに内閣府のSIP(戦略的イノベーションプログラム)を含め、厚生労働省、国土交通省、農林水産省など出口産業を所管する関係府省のプロジェクトと連携、人工知能技術の研究開発について民間投資を促進することとした。

「未来投資戦略2018—『Society 5.0』『データ駆動型社会』への変革—」(平成30年6月15日)において、第4次産業革命の新たな技術革新として、AI、ロボット、IoT及び豊富なリアルデータの活用を挙げ、これらによる様々な社会課題の解決及び大きな付加価値の創造により「Society 5.0」を実現するものとしている。

■表2-10-1 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、地図データの意味づ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。

出典：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」より作成

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みについて議論が行われた。平成28年5月に公表した中間とりまとめ<sup>\*102</sup>では、AIを産業構造を一変させうる技術として位置づけ、国費による国家プロジェクトの研究開発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(New Energy and Industrial Technology Development Organization; 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データ駆動型のAIと知識駆動型のAI

※99 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Webサイト <<http://www.nedo.go.jp/content/100790387.pdf>>

※100 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Webサイト <<http://www.nedo.go.jp/content/100862413.pdf>>

※101 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Webサイト <<http://www.nedo.go.jp/content/100862412.pdf>>

※102 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Webサイト <[http://www.meti.go.jp/committee/sankoushin/sangyougijutsu/kenkyu\\_kaihatsu\\_innovation/pdf/report01\\_01.pdf](http://www.meti.go.jp/committee/sankoushin/sangyougijutsu/kenkyu_kaihatsu_innovation/pdf/report01_01.pdf)>



の融合や計算論的神経科学の知見を取り入れた脳型AIを目指した研究開発、②様々な次世代AIのモジュール化と、それを統合するためのフレームワークの研究開発、③注力するタスクを設定し、研究成果の集約と連携のための標準的ベンチマークやデータセットの整備、⑦人工知能に関するグローバル研究拠点を活用する等による次世代人工知能の社会実装、⑧米国の卓越した研究者を招へいする等により人工知能技術開発を加速するための日米共同研究開発を実施している(表2-10-2)。また、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ション技術、及び⑥革新的な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(表2-10-3)。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<sup>\*103</sup>。さらにAI技術の社会実装促進を目的とした「次世代人工知能・ロボット技術の中核となるインテグレート技術開発」及び「人工知能技術適用によるスマート社会の実現」を平成30年度から開始している。

■表2-10-2 「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。具体的には、生活現象モデリング 地理空間情報プラットフォーム構築と空間移動のスマート化、AIを基盤としたロボット作業、科学技術研究加速のためのテキスト情報統合の4つの共有タスクを設定している。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	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、「生産性」、「健康、医療・介護」、「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。
⑧	次世代人工知能技術の日米共同研究開発(平成30年度より実施)	「人工知能の研究開発目標と産業化のロードマップ」として、取り上げるべき重点分野として特定された、「生産性」、「健康、医療・介護」、「空間の移動」の3領域および横断的な分野としての「情報セキュリティ」の領域において、喫緊に解決すべき社会課題のうち、人工知能技術による貢献が期待され、経済波及効果が見込まれる課題の解決に資する次世代人工知能技術の研究開発を人工知能技術の先進国である米国から卓越した研究者を招へいすること等により実施する。研究開発においては、産学官連携により、日本の産業競争力の強化につなげ、アウトカムの最大化を目指す。

出典:NEDO「IoT推進のための横断技術開発プロジェクト」より作成

※103 新エネルギー・産業技術総合開発機構Webサイト「IoT推進のための横断技術開発プロジェクト」<[http://www.nedo.go.jp/activities/ZZJP\\_100123.html](http://www.nedo.go.jp/activities/ZZJP_100123.html)>

■表2-10-3 「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
④	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ー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。
⑤	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ー(人工筋肉)、高度な位置制御やトルク制御を組み合わせるソフトウェア的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。
⑥	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。

出典：NEDO「IoT推進のための横断技術開発プロジェクト」より作成

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月に人工知能研究センター(Artificial Intelligence Research Center；AIRC)を設立した(表2-10-4)。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もとづくデータ駆動型のAIとWeb上の大規模な知識グラフなどにもとづく論理的・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2-10-1)。

■表2-10-4 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。
確率モデリング研究チーム	様々なデバイスから得られる実世界の大量データ(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。
機械学習研究チーム	ベイジアンモデリング、カーネル法、深層学習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な大量のデータ(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	機械学習にもと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	人の行動や身体動作の計測技術および情報共有技術により観察力を、知識構造化技術により判断力を、生活現象モデリング技術および新サービス設計技術により協働力を、各々高める各AI技術を開発し、コミュニティ内のインテリジェンス(観察、判断、行動力)を高めるための研究。
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測することを目指した研究。

(続く)

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、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。
オミクス情報研究チーム	すべての生物に共通するDNA、RNA、およびタンパク質を中心とした基盤技術、生物学における有意義な発見を目指す応用研究、バイオ実験自動化技術等に係る、バイオ産業分野に特化したAI技術を開発。
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。
データプラットフォーム研究チーム	実世界のモノ・ヒト・コトから多種多様なビッグデータの収集・蓄積・管理・利用を行うスケールなAIデータプラットフォームの実現を目指し、高精度かつ高頻度なIoT生成データを効率的に収集・格納し、利活用促進を図るためのデータガバナンス基盤技術を開発。

出典：AIST人工知能研究センターWebサイト<sup>\*104</sup>

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を整備する。これを通じて、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」<sup>\*105</sup>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、「人工知能・IoTの研究開発加速のための環境整備事業」<sup>\*106</sup>を実施した。また、深層学習の研究開発の基盤として構築した「AI橋渡しクラウド」(AI Bridging Cloud Infrastructure ; ABCI)は、2018年8月より運用を開始した。世界のスパコン速度性能ランキングTOP500 Listの5位、世界のスパコンの省エネ性能ランキングGreen 500 Listの8位を獲得している<sup>\*107</sup>。

さらに、2018年5月に、日本が取り組むべき今後のAI基盤技術の方向について、①人間と協調できるAI、②実社会で信頼できるAI、③容易に構築できるAIを提案し、意見を募集した。

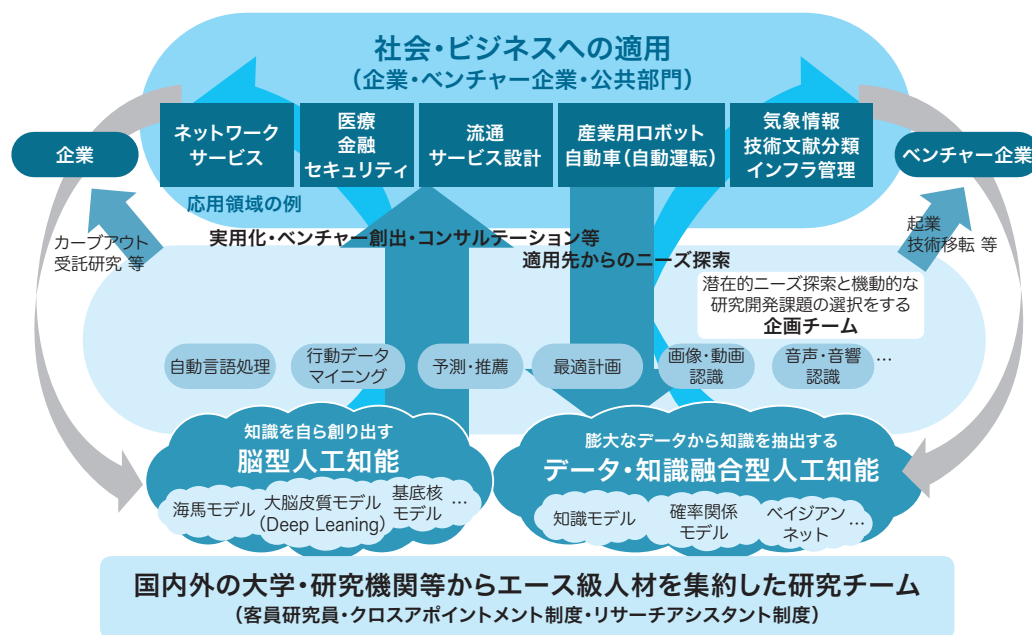
\*104 産業技術総合研究所 人工知能研究センター Webサイト<<https://www.airc.aist.go.jp/>>

\*105 「人工知能に関するグローバル研究拠点整備事業(平成28年度第2次補正予算額195億円)」経済産業省Webサイト<[http://www.meti.go.jp/information\\_2/publicoffer/282jihosei/282ji\\_PR\\_10.pdf](http://www.meti.go.jp/information_2/publicoffer/282jihosei/282ji_PR_10.pdf)>

\*106 「人工知能・IoTの研究開発加速のための環境整備事業(平成27年度補正予算額9.0億円)」経済産業省Webサイト<[http://www.meti.go.jp/information\\_2/publicoffer/27hosei/27hosei\\_PR\\_15.pdf](http://www.meti.go.jp/information_2/publicoffer/27hosei/27hosei_PR_15.pdf)>

\*107 「大規模AIクラウド計算システム「ABCI」がスパコン性能ランキング世界5位」<[https://www.aist.go.jp/aist\\_j/press\\_release/pr2018/pr20180626/pr20180626.html](https://www.aist.go.jp/aist_j/press_release/pr2018/pr20180626/pr20180626.html)>

■ 図2-10-1 AIRCにおける研究開発の取組み



出典：産業技術総合研究所Webサイト※108

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<sup>※109</sup>」を取りまとめた。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2-10-5)。

■ 表2-10-5 「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	深層学習の欠点(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う、システムに不具合が生じて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の組み合わせにより、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、さら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、さら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深層学習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。

出典：総務省「次世代人工知能推進戦略」より作成

※108 「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Webサイト <[http://www.aist.go.jp/aist\\_j/news/pr20150507.html](http://www.aist.go.jp/aist_j/news/pr20150507.html)>

※109 「次世代人工知能推進戦略」総務省Webサイト <[http://www.soumu.go.jp/main\\_content/000424360.pdf](http://www.soumu.go.jp/main_content/000424360.pdf)>

総務省所管の情報通信研究機構(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)<sup>\*110</sup>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やロボット工学の研究を実施している。さらに、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<sup>\*111</sup>するとともに、自動翻訳システムの様々な分野への対応や高精度化を進めるため、オールジャパン体制で翻訳データを集積する「翻訳バンク」の運用を開始した<sup>\*112</sup>。

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」(Advanced Integrated Intelligence Platform Project; AIPプロジェクト)<sup>\*113</sup>を推進しており、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月に設置した。当センターでは、世界最先端の研究者を糾合し、革新的な基盤技術の研究開発や我が国の強みであるビッグデータを活用した研究開発を推進することとし、具体的には表2-10-6に掲げた3つの領域で研究開発を実施することとしている。

■表2-10-6 AIPプロジェクトにおける研究領域

カテゴリ	概要
汎用基盤	深層学習の原理の解明：現在のAI技術では対応できない高度に複雑・不完全なデータ等に適用可能な基盤技術の実現。
目的指向	日本の強みを伸長：AI×再生医療・モノづくり等。 社会課題の解決：AI×高齢者ヘルスケア・防災・インフラ検査等。
倫理社会	AIと人間の関係としての倫理の明確化。 AIを活かす法制度の検討等。

出典：「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」より作成

2017年4月、AIPは、研究開発成果の実用化加速のために官業界等との連携を強化するため、東芝、NEC、富士通の3社各々との連携センターを開設し(設置期間は2022年3月31日まで)、各社が携わるソリューションを対象に、次世代人工知能基盤技術の開発から社会実装までの一貫した研究を担うこととした(表2-10-7)。

\*110 「ALAGIN 言語資源・音声資源サイト」Webサイト<<https://alaginrc.nict.go.jp/>>

\*111 先進的音声翻訳研究開発推進センター Webサイト<<http://astrec.nict.go.jp/research/index.html>>

\*112 「『翻訳バンク』の運用開始」NICTWebサイト<<https://www.nict.go.jp/press/2017/09/08-1.html>>

\*113 「『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Webサイト <[http://www.mext.go.jp/b\\_menu/houdou/28/05/1371147.htm](http://www.mext.go.jp/b_menu/houdou/28/05/1371147.htm)>

■表2-10-7 理研AIP連携センターの研究課題

連携センター名	研究課題
理研AIP-東芝連携センター	<ul style="list-style-type: none"> <li>✓ プラント生産性向上</li> <li>✓ 知的生産性向上</li> <li>✓ モビリティ自動化・ロボット化</li> </ul>
理研AIP-NEC連携センター	<ul style="list-style-type: none"> <li>✓ 少量の学習データで高精度を実現する学習技術の高度化</li> <li>✓ 未知状況での意思決定を支援する学習/AI技術の高度化</li> <li>✓ 複数AI間の調整に関わる強化学習の理論的解析</li> </ul>
理研AIP-富士通連携センター	<ul style="list-style-type: none"> <li>✓ 「ロバストな機械学習」：いかなる環境でも的確に未来を予測</li> <li>✓ 「シミュレーション・AI融合」：未知の環境の創出</li> <li>✓ 「大規模知識構造化」：より良い施策の立案</li> </ul>

出典：理化学研究所Webサイト<sup>※114</sup>より作成

科学技術振興機構 (Japan Science and Technology Agency ; JST) では、戦略的創造研究推進事業 (新技術シーズ創出) において、AIPプロジェクトに関連する研究領域をネットワークラボとして東ね、これをAIPと一体的に運営している (表2-10-8)。

■表2-10-8 ネットワークラボの構成領域

区分	研究領域	概要
CREST	共生インタラクション	人間・機械・情報環境からなる共生社会におけるインタラクションに関する理解を深め、人間同士から環境全体まで多様な形態でのインタラクションを高度に支援する情報基盤技術の創出と展開を目指す。
	人工知能	実社会の膨大なデータを知的・統合的かつセキュアに収集・処理・学習・制御するためのイノベーション創発に資する技術の確立を目指す。
	知的情報処理	人間と機械の協働により新たな知を創出し、人・集団の知的活動の質向上を実現する知的情報処理システムを目指した研究開発を推進する。
	ビッグデータ応用	個々の研究者や組織のみでは集積することが困難な大規模かつ多様な関連データを相互に関連づけて高度な統合的分析処理を行うことにより、これらのビッグデータに隠されている革新的知見や価値を抽出し創成することを実証的に研究開発する。
さきがけ	人とインタラクション	人間と人間、人間と機械、人間と情報環境、人間と実世界環境などの多様な状況でのインタラクションの進展に資する人間の能力を拡張するための新たな技術や人間と環境が高度に調和する技術の創出、インタラクション理解のさらなる深化を目指す。
	社会デザイン	モビリティなどを含めた社会基盤、介護を含むヘルスケア、防災・減災、ロボティクスなど、あらゆる分野において、情報を知的・統合的に解析・処理・制御し、新しいサービスや社会構造の構築に貢献する基盤技術を創出する。
	社会情報基盤	より良い社会の実現を目的とする情報基盤の要素技術の研究と、それらの技術を対象とする社会と調和させるために必要な制度や運用体制、ビジネスモデルまでも含めた総合的な議論と実践を行う。
CREST、さきがけ 複合領域	ビッグデータ基盤	ビッグデータの複数ドメインに共通する本質的課題を解決し、様々な分野のビッグデータの統合解析を可能にする次世代基盤技術の創出・高度化・体系化を目指す。
ACT-i	情報と未来	情報学における研究開発によって未来を切り拓く気概を持つ若手研究者を支援するとともに、新しい価値の創造につながる研究開発を推進する。

出典：JST「AIPネットワークラボの構成領域<sup>※115</sup>」より作成

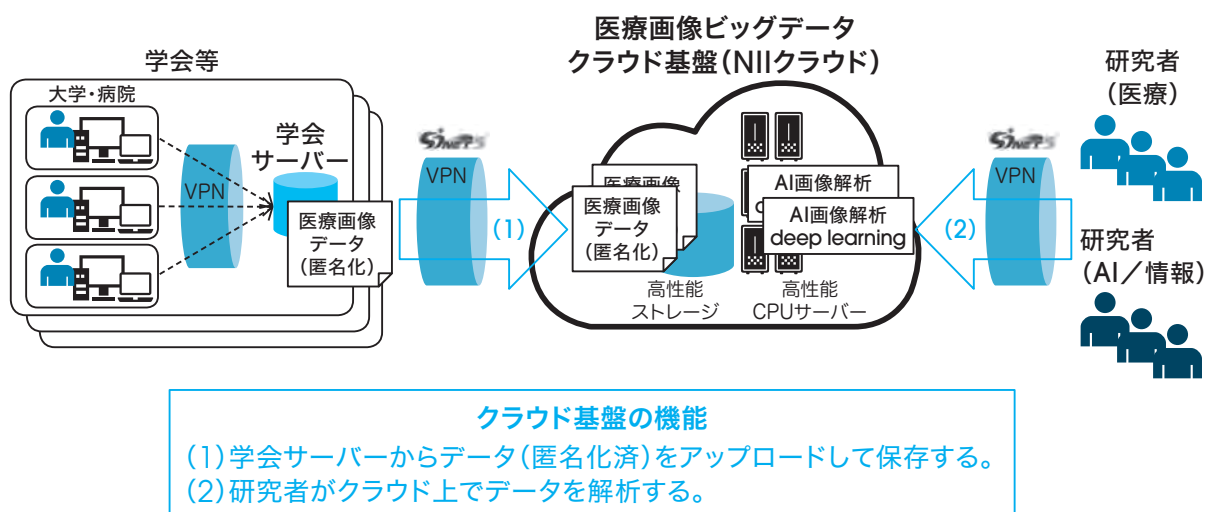
※114 「理研AIP-東芝連携センター」、「理研AIP-NEC連携センター」、及び「理研AIP-富士通連携センター」を開設 理化学研究所 Webサイト<[http://www.riken.jp/pr/topics/2017/20170310\\_1/](http://www.riken.jp/pr/topics/2017/20170310_1/)>

※115 AIPネットワークラボの構成領域<<http://www.jst.go.jp/kisoken/aip/ryoiki/index.html>>

国立情報学研究所 (NII) は、「AIが人間に取って代わる可能性がある分野は何か」といった問題を考える際の指標になりうるAIの客観的なベンチマークを指し示すことを目的として、大学入試問題をAIが解くことに挑戦した「ロボットは東大に入れるか」プロジェクトを、他機関のグループとともに2011年から推進した。2016年にはセンター試験模試で5教科8科目の合計で525点を獲得、偏差値は57.1に達し、国公立23大学、私立512大学で合格可能性80%以上との判定を得た<sup>※116</sup>。

NIIは、2017年11月に、AIをはじめネットワーク、クラウド、セキュリティなどの最先端情報技術の活用により医療分野の課題解決を推進するため「医療ビッグデータ研究センター」を設置した<sup>※117</sup>。本センターを基盤として、NIIが構築・運用する学術情報ネットワーク「SINET5」<sup>※118</sup>を活用した医療画像ビッグデータのクラウド基盤の構築、AIによる医療画像解析の研究開発、匿名化した医療画像の収集に学会<sup>※119</sup>の協力を得つつ取り組むとしている(図2-10-2)。

■ 図2-10-2 医療画像ビッグデータクラウド基盤のイメージ



出典: 国立情報学研究所ニュースリリース<sup>※120</sup>

## ②米国における国防高等研究計画局 (DARPA) 関連プロジェクト

米国政府の政策については第4章にて紹介するが、ここでは主に国立機関プロジェクトとして、DARPAのプログラムを以下に列記する(表2-10-9)。これらは、DARPAが将来必要となると予想しているAI技術であると推測されることから、米国における今後のAI技術開発の方向性を示唆するものであるといえる。特に、「説明できるAI」は政府民間を問わない共通の関心事であり、MITやGoogleなどの民間分野でも、DARPAプロジェクトとは独立に、関連研究が行われている。

※116 NIIニュースリリース<[https://www.nii.ac.jp/userimg/press\\_20161114.pdf](https://www.nii.ac.jp/userimg/press_20161114.pdf)>

※117 2017年12月25日ニュースリリース<<https://www.nii.ac.jp/news/release/2017/1225-2.html>>

※118 NIIが構築・運用している学術情報ネットワーク (Science Information NETwork)。2017年度末現在で、国立大学86校すべてを含む全国の大学や研究機関など889機関が加入しているのに加え、日米及び日欧も結ぶネットワークである。

※119 日本消化器内視鏡学会、日本病理学会、日本医学放射線学会と協力することとしている。

※120 NIIニュースリリース<<https://www.nii.ac.jp/news/release/2017/1225-2.html>>

表2-10-9 DARPAのプログラム

発表	タイトル	概要
2017年夏	「説明できるAI (Explainable AI, XAI)」	<ul style="list-style-type: none"> <li>✓ XAIは、医療・法律・金融・国防など、透明性及びユーザーから“信頼”が要求される分野で活用できるよう、意思決定の理由を人に提示できる人工知能の開発を目指す。</li> </ul>
2018年5月	L2M	<ul style="list-style-type: none"> <li>✓ 生物学システムからヒントを得て、新たな状況に継続的に対応する機械学習システムを開発する。</li> <li>✓ 2017年に公募が開始されており、2分野に取り組むチームを選出した。一つはシステム及びコンポーネントの開発を行う。もう一つは、生物有機体の学習メカニズムを研究し、これらをコンピューティングシステムに適用する。</li> </ul>
2018年7月	学習・訓練用のデータ削減	<ul style="list-style-type: none"> <li>✓ 現在の主要な機械学習システムは、人間によって個々にラベル付けされた膨大な教師データを必要とする。これが深層ニューラルネットワーク (DNN) の訓練に多くの費用と時間を要する原因となっている。また、機械学習システムは度々学習が不安定になる課題を抱えている。</li> <li>✓ そこで、機械学習システムの訓練と適合に伴う初期費用と時間を削減するため、「より少ないラベルで学習するプログラム」を開始する。</li> </ul>
2018年7月	AIE「第3の波 (Contextual Adaptation)」	<ul style="list-style-type: none"> <li>✓ AI技術領域における米国の優位性維持を目的として「人工知能探査 (Artificial Intelligence Exploration ; AIE)」プログラムを発表。</li> <li>✓ AI技術の進展を、「第1の波 (ルールベース)」、「第2の波 (統計的機械学習)」と捉え、それらの限界を認識。「第3の波」として“Contextual Adaptation”能力をもつAI、説明性やモデル (contextual model) を利用した適応性のあるAIの理論構築とアプリ開発を目指す。</li> <li>✓ 研究者は、助成受給から18カ月以内に新たなAIコンセプトの実施可能性を確立するための取組みを行う。</li> </ul>
2018年9月	常識を持つ機械	<ul style="list-style-type: none"> <li>✓ AIアプリケーションが、人とより効果的なコミュニケーションをとるための常識を有する研究開発プログラムを発表。2つのアプローチをとる。 <ol style="list-style-type: none"> <li>1. 子供のように経験から学ぶ基本的な認知の中核機能を模倣する計算モデルの開発</li> <li>2. Webリポジトリなどのテキストデータから学び、常識を形作る知識リポジトリを構築し、質問応答を行う技術の開発</li> </ol> </li> </ul>

出典: DARPA Webサイト<sup>※121</sup>より作成

### ③カナダにおけるAI政策とAIエコシステム

カナダは、第3次AIブームの火付け役となったAI研究者を次々と輩出し、「ディープラーニング (深層学習)」をはじめとするAI研究・開発の中心である。「深層学習」の理論で有名な、Geoffrey Hinton、Yann LeCun、Yoshua Bengioの3人(「カナディアン・マフィア」とも称される)が、トロント大学のカナダ先端研究機構「CIFAR」のニューラルコンピューティングプログラムに所属していたところに、「深層学習」理論を築き上げたことで知られている。この地の利を活かした戦略によりカナダはAI強国とされている。

カナダは、2017年3月にいち早くAI戦略である“Pan-Canadian Artificial Intelligence Strategy”を公表し、カナダのAI戦略<sup>※122</sup>は、①優れたAI研究者と技術力のある卒業生の数をカナダに増やす、②エドモントン、モントリオール及びトロントにあるカナダの主要な3つのAI拠点の間に科学的に卓越した連携関係を構築する、③AI促進に伴う、経済、倫理、政策及び法律についての検討にリーダーシップを発揮する、④AIに関する国家的研究コミュニティを支援する、ことを目標としている。

※121 国防高等研究計画局 (DARPA) Webサイト<<https://www.darpa.mil/>>

※122 <<https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy>>



カナダでは、深層学習の著名研究者を中心に主要大学と民間企業を結び付けた地域クラスター(センター)を中心にした発展が顕著であるとされている。まず、トロント及びオンタリオ州ウォーターloo地区は最大のAIクラスターであり、「深層学習の創始者」として世界的に有名なGeoffrey Hintonトロント大学特別名誉教授が顧問を務めるMaRSディスカバリー・ディストリクト(MaRS)というイノベーションハブを中心に、革新的技術の研究開発を支援している。MaRSには、Facebook、Google、Paypalなどが入居している。

モントリオールは2番目に大きなクラスターで、深層学習の先駆者の1人であるYoshua Bengioモントリオール大学教授が所長を務めるモントリオール学習アルゴリズム研究所(MILA)が、学術機関やインキュベーター、新興企業、既存企業などと共同で高度AI研究を進め、カナダ全域でAI技術の導入と商業化推進を目指している。

アルバータ州都のエドモントンも最近注目されているAIクラスターである。University of Alberta(アルバータ大学)はAIと機械学習の論文数で世界トップ3に入り、アルバータ・マシン・インテリジェンス研究所(AMii)には、強化学習の父として知られるRichard S. Sutton教授が在籍し、ディープラーニング(深層学習)と強化学習を組み合わせた深層強化学習の研究が進んでいる。深層強化学習を用いたAlphaGo(碁)で有名なDeepMindもエドモントンに支社を設立しアルバータ大学と緊密に協力しながら、人間が介在せずに複雑な問題に対処できるAI研究に取り組んでいる。

カナダ政府は民間のベンチャー支援の一環として、2013～2016年の4年間で4億カナダドルを出資した。AIスタートアップ企業は120社以上ある。中でも最も注目されているのは、起業家Jean-François Gagnéと共にYoshua Bengioが2016年10月に設立したAIソリューションを提供するElement AIである。同社は、2018年1月にはロンドンに欧州本社を設立、5月末時点で従業員数は270人に達し、そのうち博士号保持者は70人に達している。またGPUの次のAI開発を加速する道具とされている、「量子コンピューティング」についてもカナダには、先行するD-Wave Systemsや1QBit社等のベンチャーの存在が注目される。

このように、AI研究開発拠点を中心に、企業・人材が国内・国外から集まるとともに、ベンチャーにとって起業しやすい環境を政府が整備し、新技術を開拓するベンチャーが次々と生まれるという、エコシステムが確立しつつあるといえる。

## (2) 民間企業の研究開発の現状

AI技術は、新規産業を生む一方、幅広い産業に適用可能なものであり、AI企業と一口に言ってもその業種や業態が多様であることが特徴である。また、既存企業が活発に研究開発を行っている一方で、数多くのAIスタートアップ企業が生まれている。このプレイヤーの多様さと活発な投資活動及び企業買収活動が、人工知能技術産業のダイナミズムを生んでいるといえるだろう。

まず、AI研究開発を活発に行っている大企業についていえば、IBM Watsonを擁するIBM(米)、そしてGoogle(米)、Amazon(米)、Facebook(米)、Microsoft(米)などの巨大IT企業が挙げられる。そのほか、Samsung(韓)、NEC(日)、富士通(日)、東芝(日)、Siemens(独)といった総合電機メーカーや、NTT(日)、Qualcomm(米)、韓国電子通信研究院(韓)といった情報通信企業または研究機関、電力企業の国家电网(中)、印刷機器メーカーのXerox(米)等が挙げられる。このように、AI技術の研究開発が活発な企業の業種は多岐にわたっている。

次にAIスタートアップ企業についてみると、AIスタートアップ企業は多様かつ数多く存在するが、CB Insightsが2018“AI. 100”を発表している<sup>\*123</sup> <sup>\*124</sup>ので、このデータをもとに成

長力の高い地域と業種の傾向をみた<sup>\*125</sup>。

米国からは実に76の企業が選ばれておりAIスタートアップ企業の活動が活発であることがうかがわれる(表2-10-10)。中国からは8つの企業が選ばれているが、総資金額が米国に匹敵するほど巨大で企業あたりの規模が大きい傾向にあり多額の投資がなされている。また、総資金額が大きい業態は、ニュース及びメディア、横断的産業、フィンテック及び保険、ロボティクス、サイバーセキュリティなどである(表2-10-11)。

■表2-10-10 A.I.100企業国籍

国/地域	企業数	総資金額 (百万米ドル)
米国	76	6363.9
中国	8	5454.7
英国	5	446.7
イスラエル	4	113.6
カナダ	2	109.2
日本	2	126.2
台湾	1	81.5
フランス	1	39.7
スペイン	1	8.2
合計	100	12743.7

出典:CB Insights ai. 100 2018より作成

■表2-10-11 A.I.100 業種カテゴリ

業種カテゴリ	企業数	総資金額 (百万米ドル)
ニュース及びメディア	2	3214.1
横断的産業	10	1688.7
フィンテック及び保険	5	1399.2
ロボティクス	7	1124.4
サイバーセキュリティ	11	1064.3
ヘルスケア	8	746.6
企業内AI	8	595.5
オートテック	5	593.1
マーケティング、セールス、CRM	6	512.8
IoT	5	378.4
AIハードウェア	4	315.8
コマース	5	240.4
リスク及び規制コンプライアンス	5	175.0
生命科学	1	174.0
教育	2	138.5
地理空間分析	2	117.2
HRテック	3	68.5
農業	2	56.2
IT及び通信	1	52.9
リーガルテック	1	24.3
物理セキュリティ	2	20.9
旅行	1	14.3
ソフトウェア開発及びデバッグ	1	10.5
パーソナルアシスタンス	1	8.2
スポーツ	1	7.2
eスポーツ	1	2.7
合計	100	12743.7

出典:CB Insights ai. 100 2018より作成

\*123 CB Insights ai. 100 2018<<https://www.cbinsights.com/research-ai-100>>

\*124 These 100 Companies Are Leading the Way in A.I.<<http://fortune.com/2018/01/08/artificial-intelligence-ai-companies-invest-startups/>>

\*125 2,000以上のスタートアップ企業の中から、投資指数、技術革新、チーム力、特許活動、モザイクスコア(CB Insights独自の指数)、資金調達歴、評価、ビジネスモデルを含む基準で選ばれている。

2016年9月にFacebook、Amazon、Alphabet (Google)、IBM、Microsoftの5社は、共同でAIにおける新たな提携を行い、非営利団体“Partnership on AI”を設立した<sup>\*126</sup>。AI技術について、その研究開発に関するベストプラクティス構築について支援するとともに、普及活動等を行うことを目的としている。そして、安全が求められる重要用途AIや公正で透過的で説明可能なAIの実現、AIによる労働及び経済への効果、人と協調できるAIシステム、AIが社会に与える影響、AIと社会的な善について検討を行っている。現在、9カ国50以上の企業及び団体が参加している<sup>\*127</sup>。

### 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みについて現状では米国の情報系企業を中心とする取組みに比べスタートが遅れたことは否めない。そもそも国内のAI関連の人材は不足しており(「2.7.1 AI人材育成の全体イメージ」参照)、大手企業は米国への投資や直接研究所を開設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。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、トヨタ自動車及びファナックはPreferred Networksに、ソフトバンクはFlipkart(印)、Fanatics(米)等に、NVIDIAはABEJA等に、各々出資している。

NTTグループは、AI関連技術群の総称としてcorevoをブランド名としたサービスを展開している。corevoを構成するAI技術は、人の発する情報を捉えて意図・感情を理解するAgent-AI(コンタクトセンターや高齢者支援)、心身を読み書き深層心理・知性・本能を理解するHeart-Touching-AI(スポーツ上達やメンタルウェルネス)、人・モノ・環境を読み書き、瞬時に予測・制御するAmbient-AI(運転支援や災害予測・復旧)、複数のAIがつながり社会システム全体を最適化するNetwork-AI(ネットワークの故障予知や社会最適化)の4種からなる<sup>\*128 \*129</sup>。

他の特許出願数上位企業としてNEC、富士通、東芝が挙げられ、既存サービスにAI技術を導入した製品展開がなされている。NECは、デジタルトランスフォーメーションを加速する最先端AI技術「NEC the WISE」を展開している<sup>\*130</sup>。富士通は、AIを活用したサービスZinraiを展開しており<sup>\*131</sup>、また、物質の構造など実世界のデータの関係(グラフ構造)を直接学習する「Deep Tensor」と学術論文など世界中に存在する膨大な知識を構造化したナレッジグラフを組み合わせAIが行った判断結果にいたる根拠を提示する技術を開発した<sup>\*132</sup>。東芝は、製造現場等に適用するAI技術SATLYSと人と人のコミュニケーションをサポートするAI技術RECAIUSを展開している<sup>\*133</sup>。

Preferred Networksは、Preferred Infrastructureからスピナウトして設立されたスタートアップ企業であり、リアルタイム機械学習技術をIoT領域に適用している。事例として、交通シス

---

\*126 <<https://www.partnershiponai.org/>>

\*127 2018年7月20日現在。

\*128 NTTグループのAI技術「corevoTM」を実装した共通基盤にてデバイス連携システムの合同実証実験を開始  
<<http://www.ntt.co.jp/news/2016/1607/160725a.html>>

\*129 研究Q&A <<http://www.ntt.co.jp/svlab/activity/pickup/qa47.html>>

\*130 デジタルトランスフォーメーションを加速する最先端AI技術群「NEC the WISE」 <<https://jpn.nec.com/ai/>>

\*131 Zinraiトップ <<http://www.fujitsu.com/jp/solutions/business-technology/ai/ai-zinrai/>>

\*132 AIの推定理由や根拠を説明する技術を開発「Deep Tensor」とナレッジグラフを融合  
<<http://pr.fujitsu.com/jp/news/2017/09/20-1.html>>

\*133 人とモノが共働する世界を実現する東芝の2つのAI技術 <<https://www.toshiba.co.jp/iot/spinex/ai.htm>>

テム(2014年10月からトヨタ自動車と共同研究)、製造業(2015年6月からファナックと、2017年12月から日立製作所と共同研究)、バイオヘルスケア(2017年12月から国立がん研究センター棟と共同研究)を重点事業領域としている。同社はニューラルネットワークのフレームワークChainerを公開しているのに加え、海外学会においても積極的に発表を行っており、高く評価されている。LeapMindは、企業向けにディープラーニングエッジソフトウェア開発を行い、また、深層学習を実装したエッジデバイスの技術開発を行っている。

2017年6月に日本ディープラーニング協会<sup>\*134</sup>が設立され、ディープラーニングについて、①産業活動の促進、②公的機関や産業界への提言活動、人材育成、国際連携活動、社会との対話、等の活動を行っていくとしている。同協会は、ディープラーニングに関する知識普及と人材育成のため、ジェネラリスト向けのG検定及びエンジニア向けのE検定の2種類の資格試験を実施している(「2.7.2 スキル標準/認定・検定制度」参照)。

我が国では、昨今のAIの興隆への対応は少し遅れたものの、今後に関しては、AI及び脳科学等関連諸科学のアカデミックの研究者層の厚みを背景とした産官学連携の推進、リアル空間のデータを持つ製造業の強みを利用したビジネス開発など、既存の強みを活かした戦略が期待される。

## 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Geoffrey Hintonは、トロント大学の教授であるが、現在はAlphabet(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。Hintonのほかにも、Facebook AI ResearchのYann LeCun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。

Jeffrey Deanが率いるGoogle AI<sup>\*135</sup>は、1,700名弱(2018年7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、深層学習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。

Amazonは、2017年中に米国企業中最大の230億ドル近い研究開発予算を投じ、これは2位のAlphabet(Google)を大きく引き離していることで話題となった<sup>\*136</sup>。同社は、AWS、Alexa、キャッシュレスのコンビニエンスストアAmazon Goといった野心的な技術に投資を行っている。なお、2017年の米国企業R&D投資額ランキング3位以降は、順にIntel、Microsoft、Appleであり、実に上位5位までをIT企業が占めている。

Facebookも、AI技術に投資を行っており、35億件もの一般入手可能な写真のデータセットに対し、ハッシュタグを付与する画像認識システムの訓練に成功したと発表した。また、オープンソースAIフレームワークPyTorchを公開している<sup>\*137</sup>。

※134 一般社団法人日本ディープラーニング協会Webサイト<<http://www.jdla.org/>>

※135 2018年5月にGoogle Researchから改称されたことが発表された。

※136 Amazon spent nearly \$23 billion on R&D last year — more than any other U.S. company <<https://www.recode.net/2018/4/9/17204004/amazon-research-development-rd>>

※137 開発者カンファレンス「F8」2018：2日目発表内容まとめ <<https://ja.newsroom.fb.com/news/2018/05/f8-2018-day2/>>

GAFAsの残る一角であるAppleは、これまで(Siriを提供していたものの)人工知能などの知的ソフトウェアの分野で立ち遅れていると目されていたが、Googleの人工知能及び検索の責任者であったJohn Giannandreaを迎えたこと<sup>\*138 \*139</sup>から分かるように、AI技術分野を強化しているところである。

注目すべき点として、Amazon、Google、Microsoft、Appleは、スマートスピーカーを発売しており、各々特徴は異なるものの、音声認識技術、会話技術、パーソナルアシスタンス技術、スマートホーム技術を家庭に浸透させつつある。

IBMは、「Watson」を含むクラウドサービスを提供する企業である。「Watson」はIBMが研究開発する人間の認知に関わる情報処理(コグニティブ・コンピューティング)の総称であり、同社はWatsonを開発するプロジェクトに2006年から取り組んでいる。そして、IBMとMITは、2017年9月にMIT-IBM Watson AI labを設立するために10年間で2億4,000万米ドルに及ぶ契約に合意したと発表した。この研究所では、まず第1に、ニューラルネットワークにもとづく深層学習を用い、特定応用をさらに推し進めるAIアルゴリズムを開発、第2に機械学習と量子コンピューティングを結び付ける研究を行うとしている<sup>\*140</sup>。

アレン人工知能研究所(Allen Institute for Artificial Intelligence; AI2)は、Microsoftの共同創業者であるポール・アレンの出資を受け、人工知能に関する重大な問題を探求することを目的として、2014年に設立された非営利の研究所である。同研究所が推進する“common sense AI”プロジェクトは、ポール・アレンから新たに1億2,500万米ドルの出資を受けており、人工知能に常識を持たせることで大きなブレイクスルーをもたらすことを目指している<sup>\*141</sup>。DARPAの常識を持つ機械を目指すプロジェクトに対して、ベンチマークテストを提供することになっている。

Upstartは、貸付プロセスの自動化にAI技術を適用するスタートアップ企業であって、顧客の学歴や就労情報から将来の収入を予想しリスクを算出して貸し付けを行う。同社は銀行やクレジット会社等に当該サービスをSaaSとして提供している。

Cybereasonは、イスラエル軍でサイバーセキュリティに携わった共同創業者により創業されたスタートアップ企業であり、エンドポイントのログを収集してリアルタイムで監視して解析を行い、攻撃者の侵入を検知するプラットフォームを提供している。日本法人サイバーリーズン・ジャパン株式会社は、ソフトバンクとCybereasonの合弁会社である。

### 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AI研究が盛んであるとともに、DeepMindの本拠地でもある。深層学習と強化学習の組み合わ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(Recurrent Neural Network; RNN)、強化学習の一種であるQ学習(Q-learning)を用いたシステムの開発など、今後のAI開発の方向性を

---

\*138 Apple、GoogleのAIトップを引き抜く<<https://jp.techcrunch.com/2018/04/04/2018-04-03-apple-steals-googles-ai-chief/>>

\*139 ニューヨークタイムズ紙、2018年4月3日<<https://www.nytimes.com/2018/04/03/business/apple-hires-googles-ai-chief.html>>

\*140 IBMとMITがAI研究パートナーシップを締結、10年間で2億4000万円が提供される<<https://jp.techcrunch.com/2017/09/07/20170906ibm-and-mit-pen-10-year-240m-ai-research-partnership/>>

\*141 アレン人工知能研究所<<https://allenai.org/>>

見据えた取組みを多く実施している。

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。

ドイツのシーメンスは、2017年9月に中国の精華大学と協力して北京に自律ロボットの研究開発センターを設立すると発表した<sup>\*142</sup>。これは、国際ロボット連盟(IFR)のレポート<sup>\*143</sup>に、2019年には中国が世界最大のロボット市場になると見込まれたことを受けたものと考えられる。

Shift Technologyは、フランスのパリで設立されたスタートアップ企業であり、損害保険や生命保険などの保険金詐欺検出にAIを利用したソリューションを提供している。

Babylon Healthは、デジタルヘルスケアアプリケーションを提供する英国のスタートアップ企業である。ユーザーはこのアプリケーションを通じて一般開業医とビデオ面談をすることができる。後継バージョンでは、医師とのビデオ面談に先立ってAIによる医療アドバイスを受けることができる。このチャットボットを使うことで、不必要な通院を減らす効果がある。

#### 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(3社を合わせて“BAT”と呼ばれることがある)等の情報系企業のAIへの取組みが先行している。

Baidu(百度)は、2000年に創業された中国最大の検索エンジンを提供する企業であり、全世界の検索エンジン市場においてGoogleに次いで第2位、中国国内では最大のシェアを占める。2013年に北京に、2014年にはシリコンバレーに人工知能研究所を開設し、3億ドルを投じている。同社は、検索サービスの向上につながる画像認識や音声認識、自然言語解析、機械翻訳などの研究に取り組んでいる。

Alibabaは、1999年の創業で、電子商取引サイトを運営している。ディープラーニングと推薦システム、マルチメディアデータ分析等の研究を行っている。

Tencentは、1998年に創業されたインターネット企業であり、ゲームの売上高では世界最大である。同社は、機械学習、コンピュータービジョン、音声認識、自然言語処理の研究を行っている。

国家電網は、世界最大の電力配送会社である。中国国内での特許取得件数で首位であり<sup>\*144</sup>、人工知能技術分野でも出願数で2位に大きく差をつけて首位になっている。

Bytedanceは、ニュースアプリで成長したスタートアップ企業であり、総資産額は31億米ドルを超えている。データマイニング技術とリコメンド技術をニュースアプリに応用している。

Face++は、顔認識クラウドサービスプラットフォームを構築するスタートアップ企業であり、顔認証サービスの提供に加えて顔決済サービスを実用化している。

### (3) 特許・論文の動向

#### 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

※142 シーメンス中国が国際自律ロボット開発を主導 <[http://jp.xinhuanet.com/2017-09/15/c\\_136611930.htm](http://jp.xinhuanet.com/2017-09/15/c_136611930.htm)>

※143 The Impact of Robots on Productivity, Employment and Jobs <[https://ifr.org/img/office/IFR\\_The\\_Impact\\_of\\_Robots\\_on\\_Employment.pdf](https://ifr.org/img/office/IFR_The_Impact_of_Robots_on_Employment.pdf)>

※144 2017年の特許取得件数において国家電網が首位に <[http://www.spc.jst.go.jp/news/180105/topic\\_3\\_03.html](http://www.spc.jst.go.jp/news/180105/topic_3_03.html)>

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。

以下に、AIに関わる特許の動向を示す。調査対象は、優先権主張年が2000年から2018年の期間の特許・実用新案出願<sup>\*145</sup>（以下、この項では「特許出願」という）とし、データベースにクラリベイト・アナリティクスのDerwent Innovationを使用し、国際特許分類（International Patent Classification；IPC）の分類コード（「G06N（特定の計算モデルにもとづくコンピューター・システム）」及びその他の人工知能技術関連分類<sup>\*146</sup>）を用いて検索を行った<sup>\*147</sup>。

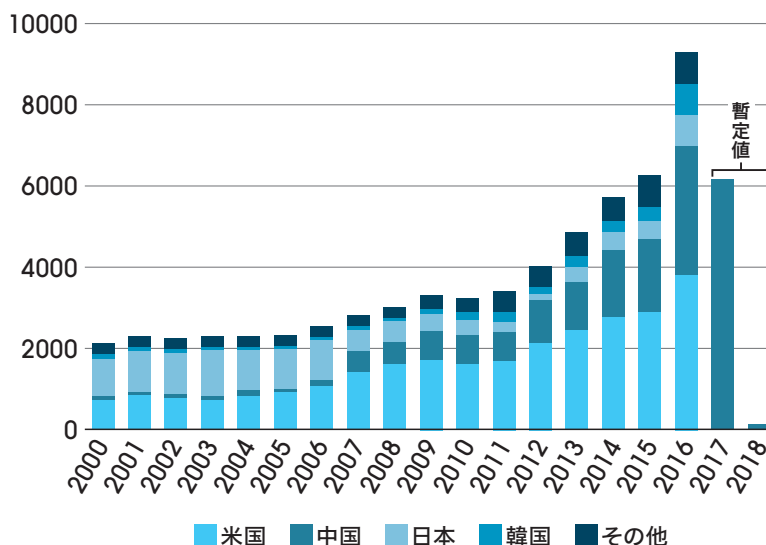
表2-10-12は、2000年から2018年の間の特許出願累計についての出願受理機関ごとのランキングである<sup>\*148</sup>。図2-10-3は、特許出願総数の推移であり<sup>\*149</sup>、2002年から2005年までは横ばい、2006年から2009年までは増加傾向にあるものの2010年に一旦減少したあと、2011年以降は急増している。そして、2002年ごろは日本と米国が大半を占めていたものの、次第に日本の相対的割合が低下しつつあることがみてとれる。米国及び中国の特許出願数は大きく伸びており、特に2016年以降は、他国出願の大半は公報発行前であることに留意は必要であるが、中国出願数の著しい伸びが継続している点は注目すべきところである。

■表2-10-12 2000年～2018年の国／地域ごとのAI関連特許出願数

順位	国／地域	出願数	出願割合
1	米国	31,055	44.5%
2	中国	18,132	26.0%
3	日本	9,866	14.1%
4	韓国	3,059	4.4%
5	国際	1,358	1.9%
6	欧州	1,100	1.6%
7	台湾	973	1.4%
8	独国	815	1.2%
9	英国	663	0.9%
10	インド	562	0.8%
11	ロシア	532	0.8%
12	フランス	489	0.7%
13	その他	1,234	1.8%
	合計	69,838	100.0%

出典：Derwent Innovationでの検索結果をもとにNEDO技術戦略センターが作成（2018）

■図2-10-3 世界のAI関連特許出願数推移



出典：Derwent Innovationでの検索結果をもとにNEDO技術戦略センターが作成（2018）

※145 パテントファミリーが存在する場合、優先権主張年が最先の出願（ベーシック出願）をカウントした。パテントファミリーが存在しない出願は出願年を用いてカウントした。また、出願日から公報発行まで期間が設けられているため（多くの場合18カ月）、検索日との関係から2017年ないし2018年は暫定値であり、それ以前の値もデータベースの更新等により変動しうる。

※146 G06Nのほか、G06F15/18、G06F17/28、G06F19/24、G06G7/60、G06K9/66、G10L15/06、G10L15/065、G10L15/07、G10L15/14、G10L15/16、G10L15/18、G10L15/183、G10L15/187、G10L15/19、G10L15/193、G10L15/197、G10L17/04、G10L17/18、G10L25/30、G10L25/33、G10L25/36、G10L25/39を対象とした。

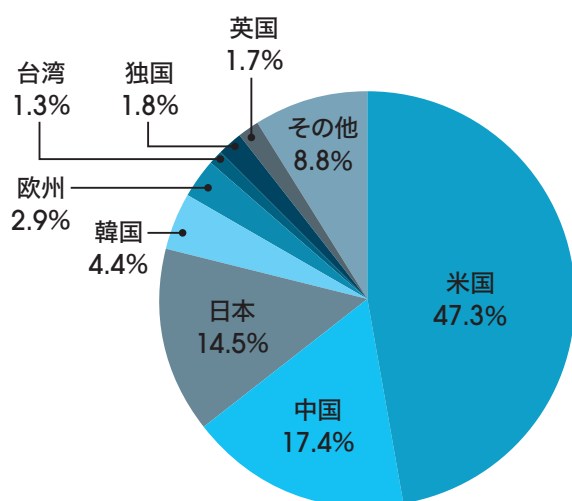
※147 検索日は2018年6月28日。

※148 表中の「国際」は世界知的所有権機関（World Intellectual Property Organization；WIPO）に、「欧州」は欧州特許庁（European Patent Office；EPO）に出願され発行されたものである。

※149 出願から特許調査データベース更新までにタイムラグがあるため、2017～2018年は暫定値である。また、2016年以前であっても、特許調査データベースの更新により若干の変動がありうる。

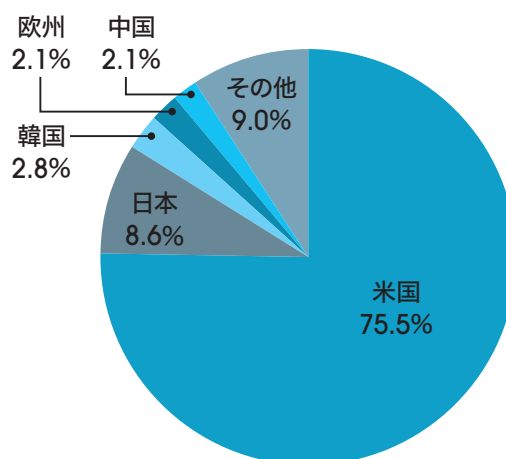
次に、国ごとの出願傾向を示す。対象出願は優先権主張年が2000年から2018年の各国に対する特許出願<sup>※150</sup>である。図2-10-4の世界特許出願の出願人国籍<sup>※151</sup>をみると、米国が半数近くを占めており、次いで中国、日本、韓国の順となっている。図2-10-5～図2-10-8は、各国における出願人国籍である。一般に通常は当該国の出願人が多くの割合を占めるため、注目すべきは2位以降の国であり、どの国の企業が当該国を、競争力を維持すべき市場として重視しているかを反映していると考えられる。図2-10-5の米国への出願の場合、日本、韓国、欧州の順で出願がなされている。図2-10-6の中国への出願の場合、米国、日本の順となっている。図2-10-7の日本への出願の場合、米国、欧州の順となっている。図2-10-8の独国は少し状況が異なり、同国の出願数が大きくないこともあって外国籍の比率が高く、米国と独国がほぼ同程度であり、日本、欧州が続く。

■ 図2-10-4 世界におけるAI関連特許出願の出願人国籍 (2000年～2018年累計)



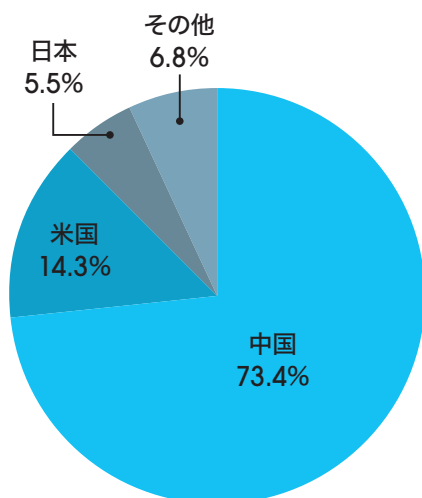
出典: Derwent Innovationでの検索結果をもとに NEDO技術戦略センターが作成(2018)

■ 図2-10-5 米国出願(AI関連)の出願人国籍 (2000年～2018年累計)



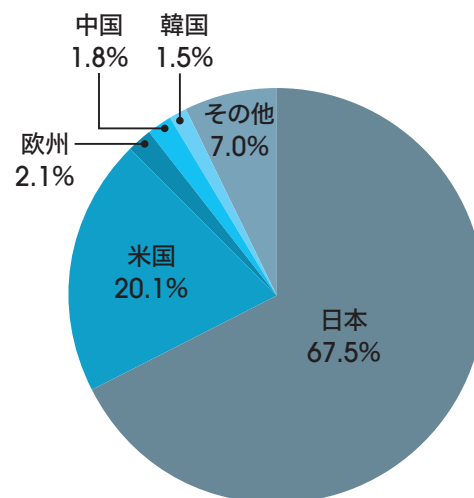
出典: Derwent Innovationでの検索結果をもとに NEDO技術戦略センターが作成(2018)

■ 図2-10-6 中国出願(AI関連)の出願人国籍 (2000年～2018年累計)



出典: Derwent Innovationでの検索結果をもとに NEDO技術戦略センターが作成(2018)

■ 図2-10-7 日本出願(AI関連)の出願人国籍 (2000年～2018年累計)

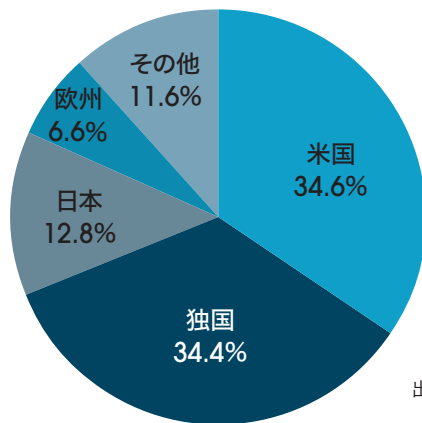


出典: Derwent Innovationでの検索結果をもとに NEDO技術戦略センターが作成(2018)

※150 各国に対する特許出願をカウントした。ただし、ある国に同一の優先権主張を伴う出願が複数件ある場合は1件とカウントした。  
 ※151 ファミリー中の最先の出願(ベーシック出願)の出願先の国を出願人の国籍とみなした。



■ 図2-10-8 独国出願 (AI関連) の出願人国籍 (2000年～2018年累計)



出典: Derwent Innovationでの検索結果をもとに  
NEDO技術戦略センターが作成(2018)

次に、世界全体において2012年から2016年までの出願累計が多い出願人を示す(表2-10-13)。IBMが首位であり、出願数が他に比べて多い。次いで、Google、Microsoftと米国企業が続き、韓国企業のSamsung、日本企業のNTT、NECが続く。トップ15に入った出願人国籍は、米国、日本、中国、韓国、独国となっている。

■ 表2-10-13 2012年～2016年AI関連特許出願の多い出願人ランキング

順位	2012～2016年特許出願数トップ15	出願数
1	IBM(米)	2396
2	Google(米)	816
3	Microsoft(米)	718
4	Samsung(韓)	432
5	NTT(日)	380
5	NEC(日)	380
7	国家电网(中)	353
8	富士通(日)	267
9	Facebook(米)	247
9	Qualcomm(米)	247
11	韓国電子通信研究院(韓)	240
12	東芝(日)	194
13	Amazon(米)	186
14	Siemens(独)	172
14	Xerox(米)	172

出典: Derwent Innovationでの検索結果をもとにNEDO技術戦略センターが作成(2018)

次に、2016年における、世界全体と米国、中国、日本、独国における出願数が多い出願人を示す(表2-10-14)。米国においては、世界全体の傾向がおおむね反映されている。IBMが首位であり、Microsoftが追う状況にある。中国では、中国全土へ送電・変電・配電を行う国家电网が一位、IT企業大手のAlibabaが2位、そのほか中国の企業及び大学が大半を占め、その中でGoogle(米)やSamsung(韓)といった外国企業がランキングに入っているところから、中国国内では大規模出願人により出願されているというよりは、多数の出願人によって全体として大量の出願がなされているものと推測される。日本では、日本の大手総合電機メーカーなど日本の企

業及び研究機関が国内出願の上位を独占している状況にある。独国は米国企業、独国企業、日本企業が含まれるが、独国の産業構造を反映して自動車業界の企業が比較的多いことが特徴である。

表2-10-14を図2-10-5と併せてみると、米国市場に対して、米国企業のみならず日本、韓国、独国の大手外国企業が出願している状況がみられ、グローバル企業の米国市場重視の傾向をみる事ができる。また、図2-10-5～図2-10-8をみると、中国の企業及び大学の外国での特許出願活動は今のところさほど活発化していないが、現状は中国国内の出願件数の伸びが著しい状況にあることから、今後の動向には注視していく必要がある。

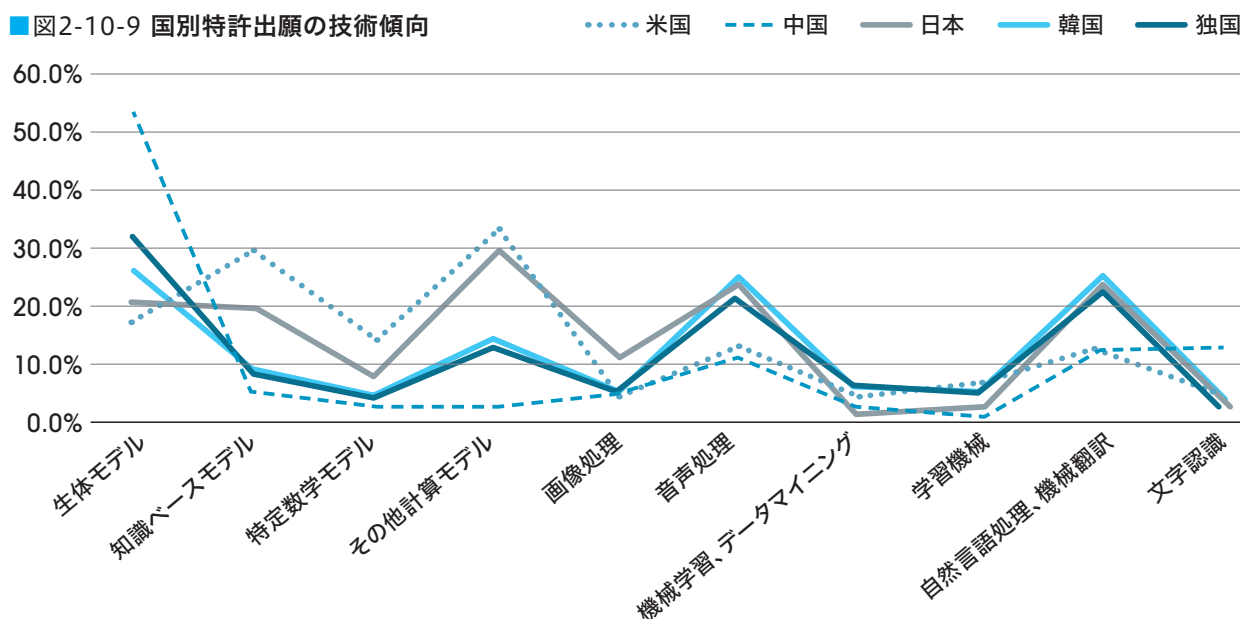
■表2-10-14 2016年における国ごとの上位出願人による特許出願数(AI関連)

順位	世界		米国		中国		日本		独国	
	出願人	出願数	出願人	出願数	出願人	出願数	出願人	出願数	出願人	出願数
1	IBM(米)	666	IBM(米)	666	国家電網(中)	121	NTT(日)	95	Google(米)	56
2	Microsoft(米)	192	Microsoft(米)	188	Alibaba(中)	56	富士通(日)	92	ファナック(日)	33
3	Google(米)	189	Google(米)	183	西安電子科技大(中)	51	パナソニック(日)	49	フォード(米)	27
4	Samsung(韓)	126	Samsung(韓)	111	Huawei(中)	48	東芝(日)	47	Bosch(独)	12
5	国家電網(中)	121	Facebook(米)	83	Google(米)	47	ファナック(日)	35	ゼネラルモーターズ(米)	12
6	富士通(日)	102	富士通(日)	81	Baidu(中)	46	日本電気(日)	35	アドビシステムズ(米)	8
7	NEC(日)	102	NEC(日)	69	アモイ大(中)	42	日立(日)	33	Siemens(独)	8
8	NTT(日)	95	Intel(米)	67	天津大(中)	41	SONY(日)	30	Audi(独)	4
9	韓国電子通信研究院(韓)	85	アクセンチュア(アイルランド)	56	北京工業大学(中)	41	三菱電機(日)	27	BSH(独)	3
10	Facebook(米)	83	LinkedIn(米)	48	北京バイドゥ・ネットコム(中)	40	トヨタ自動車(日)	26	Lenovo(中)	3
11	Intel(米)	73	Siemens(独)	48	華南理工大(中)	40	ヤフー(日)	25	NVIDIA(米)	3
12	Alibaba(中)	60	アドビシステムズ(米)	44	浙江大(中)	38	キヤノン(日)	21	トヨタ自動車(日)	3
13	アクセンチュア(アイルランド)	58	東芝(日)	44	河海大(中)	37	富士ゼロックス(日)	19	Apple(米)	2
14	Siemens(独)	57	Apple(米)	39	Samsung(韓)	36	情報通信研究機構(日)	15	コグネックス(米)	2
15	パナソニック(日)	55	GE(米)	39	南京郵電大(中)	33	デンソー(日)	15	DeepMind(英)	2

出典: Derwent Innovationでの検索結果をもとにNEDO技術戦略センターが作成(2018)

次に、特許出願数上位国の米国、中国、日本、韓国に独国を加えた5カ国について、特許出願が属する技術分野<sup>※152</sup>を図2-10-9に示した。特定の国の特許出願において各々の技術分野の出願が占める割合を示している。技術分野の内容は表2-10-15に記載した。

中国特許出願は、ニューラルネットワーク関連の技術分野(画像、音声、自然言語処理等)の割合が相対的に高く、中国では深層学習等のニューラルネットワークモデルに特化した技術開発がなされている。一方、米国特許出願及び日本特許出願は、知識ベースモデル及びその他計算モデルに関する特許出願の割合が相対的に高く、よりバラエティに富む研究開発がなされていると考えられる。



出典: Derwent Innovationでの検索結果をもとにNEDO技術戦略センターが作成(2018)

■ 表2-10-15 技術分類

技術	技術の概要
生物モデル	生物モデルにもとづくコンピューターシステム 例: ニューラルネットワーク、深層学習、遺伝的モデル
知識ベースモデル	知識ベースモデルを利用したコンピューターシステム 例: 知識の表現、推論
特定数学モデル	特定の数学的モデルにもとづいたコンピューターシステム 例: ファジー論理、カオスモデル、非線形モデル
その他の計算モデル	特定の計算モデルにもとづくコンピューターシステムであって、生物モデル、知識ベースモデル、特定数学モデルのいずれにも分類されない主題事項に関するもの
画像処理	汎用イメージ処理 例: ニューラルネットワークを用いた画像認識
音声処理	汎用音声処理 例: ニューラルネットワークを用いた音声認識
機械学習、データマイニング	機械学習、データマイニング等を用いた特定用途コンピューター
学習機械	自身が得た経験に応じてプログラムが変化するコンピューター
自然言語処理、機械翻訳	自然言語の処理または翻訳を行うコンピューター
文字認識	手書き文字認識、指紋等のパターン認識

出典: IPC分類表(特許庁,2018)をもとにNEDO技術戦略センターが作成(2018)

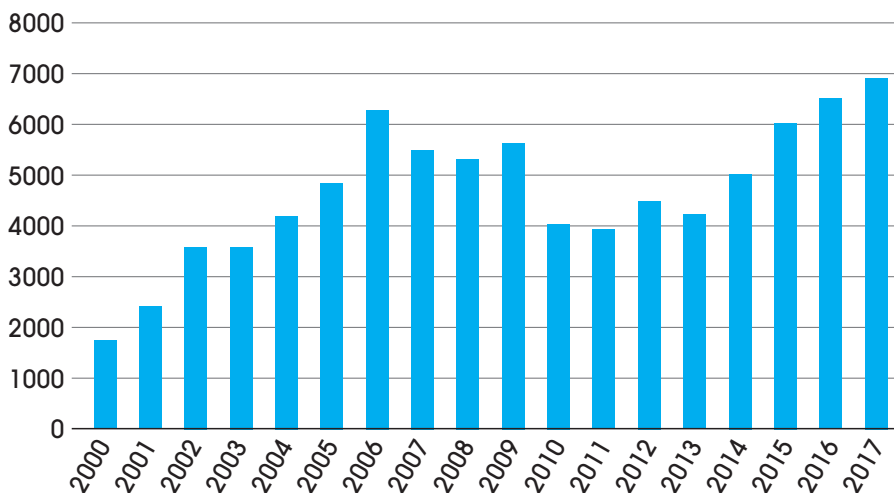
※152 2012年から2016年の特許出願について、パテントファミリーに付与された特許分類についてベーシック出願の出願先の国ごとに整理し、技術グループ単位の付与数をカウントしたものをパテントファミリー数で除して割合とした。

## ②論文動向

下図に、機械学習等の人工知能技術に関する論文について、2000年～2017年の論文発行数の推移を示す<sup>※153</sup>。2006年まで増大した後に一旦減少に転じ、2013年以降は増加の一途をたどっていることが分かる(図2-10-10)。

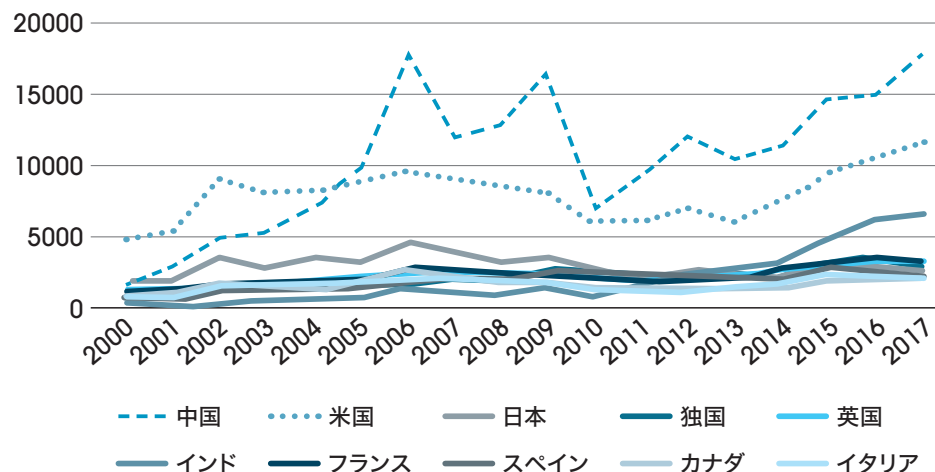
2013年以降について国別に見ると、主要国はいずれも論文発行数が増加しており、論文発行数の順位は、中国、米国、インド、英国、独国と続き、日本は6位であった(図2-10-11)。

■ 図2-10-10 世界の人工知能技術論文等発行数(2000年～2017年)



出典: Web of Science Core Collectionでの検索結果をもとにNEDO技術戦略センターが作成(2018)

■ 図2-10-11 著者所属機関の国/地域別人工知能技術論文等発行数



出典: Web of Science Core Collectionでの検索結果をもとにNEDO技術戦略センターが作成(2018)

※153 論文の検索にはクラリベイト・アナリティクスのWeb of ScienceのCore Collectionを使用した。検索キーワードはTS= (“artificial intelligence”) OR TS= (“neural network\*” AND computer\*) OR TS= (“deep learning”) OR TS= (“machine learning”) OR WC= (“artificial intelligence”)とした。対象国は特に限定していない。検索日は2018年8月3日～8月16日。

2013年から2018年に発行された人工知能技術に関する論文30万9,388件に対して、キーワード検索<sup>\*154</sup>を行ったときのヒット論文数のシェアを示す(表2-10-16)。機械学習及びニューラルネットワークについては、中国、米国の順であるが、シェアは同程度である。また、自然言語処理及び音声認識は米国が優勢である。日本は、音声認識及びロボティクスが比較的健闘している。

■表2-10-16 国/地域別キーワードヒット論文シェア(2013年～2018年)

順位	全体		機械学習 ニューラル ネットワーク		自然言語処理		画像認識		音声認識		ロボティクス	
	国	シェア	国	シェア	国	シェア	国	シェア	国	シェア	国	シェア
1	中国	24.2%	中国	23.7%	米国	23.3%	中国	27.0%	米国	24.3%	中国	15.0%
2	米国	16.1%	米国	22.4%	中国	14.1%	米国	21.5%	中国	13.3%	米国	14.9%
3	インド	8.1%	インド	7.7%	インド	8.5%	インド	5.9%	インド	7.9%	独国	8.2%
4	英国	5.0%	英国	5.8%	英国	5.4%	英国	5.5%	日本	7.7%	日本	8.2%
5	独国	4.9%	独国	4.5%	スペイン	5.2%	独国	4.6%	独国	6.6%	仏国	5.1%
6	日本	4.7%	日本	4.0%	イタリア	4.5%	スペイン	4.5%	英国	5.7%	イタリア	5.0%
7	仏国	4.3%	カナダ	3.7%	独国	4.4%	オーストラリア	4.4%	仏国	4.8%	英国	5.0%
8	スペイン	4.0%	スペイン	3.7%	仏国	4.1%	仏国	4.1%	スペイン	3.9%	韓国	4.6%
9	イタリア	3.5%	オーストラリア	3.6%	日本	3.7%	イタリア	3.9%	イタリア	3.4%	スペイン	4.0%
10	カナダ	3.1%	仏国	3.2%	カナダ	3.4%	日本	3.7%	シンガポール	2.8%	インド	3.8%

出典: Web of Science Core Collectionでの検索結果をもとにNEDO技術戦略センターが作成(2018)

※154 「項目名(ヒット件数): 検索式」は次のとおりである。機械学習+ニューラルネットワーク(95,457): TS=("NEURAL NETWORK\*") OR TS=("DEEP LEARNING") OR TS=("LEARNING MACHINE\*") OR TS=("MACHINE LEARNING") OR TS=(SVM) OR TS=(SUPPORT VECTOR MACHINE\*) OR TS=("REGRESSION\*")、自然言語処理(4,552): TS=("NATURAL LANGUAGE\*")、画像認識(8,908): TS=("IMAGE RECOGN\*") OR TS=("VISION RECOGN\*") OR TS=("COMPUTER VISION\*")、音声認識(2,756): TS=("SPEECH RECOGN\*")、ロボティクス(38,580): WC=("ROBOT\*") OR TS=("ROBOT\*")

## ③研究拠点

人工知能技術分野において論文発表数の多い研究機関を表2-10-17にまとめた。

■表2-10-17 人工知能技術論文数が多い研究機関一覧(2013～2018年)

順位	機関	分類(国/地域)	論文数
1	CHINESE ACADEMY OF SCIENCES	大学(中国)	5999
2	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS	国立研究機関(仏国)	4997
3	UNIVERSITY OF CALIFORNIA SYSTEM	大学(米国)	3935
4	INDIAN INSTITUTE OF TECHNOLOGY SYSTEM IIT SYSTEM	大学(インド)	2954
5	TSINGHUA UNIVERSITY	大学(中国)	2585
6	HARBIN INSTITUTE OF TECHNOLOGY	大学(中国)	2216
7	UNIVERSITY OF LONDON	大学(英国)	2211
8	NANYANG TECHNOLOGICAL UNIVERSITY	大学(中国)	2167
8	NANYANG TECHNOLOGICAL UNIVERSITY NATIONAL INSTITUTE OF EDUCATION NIE SINGAPORE	大学(シンガポール)	2167
10	INRIA	国立研究機関(仏国)	2133
11	UNIVERSITY OF TEXAS SYSTEM	大学(米国)	2093
12	CARNEGIE MELLON UNIVERSITY	大学(米国)	2028
13	BEIHANG UNIVERSITY	大学(中国)	1907
14	UNIVERSITE PARIS SACLAY COMUE	大学(仏国)	1893
15	SHANGHAI JIAO TONG UNIVERSITY	大学(中国)	1733
16	STATE UNIVERSITY SYSTEM OF FLORIDA	大学(米国)	1732
17	ISLAMIC AZAD UNIVERSITY	大学(イラン)	1671
18	MASSACHUSETTS INSTITUTE OF TECHNOLOGY MIT	大学(米国)	1611
19	ZHEJIANG UNIVERSITY	大学(中国)	1552
20	UNIVERSITY OF CHINESE ACADEMY OF SCIENCES CAS	大学(中国)	1528
21	NATIONAL UNIVERSITY OF SINGAPORE	大学(シンガポール)	1526
22	HUAZHONG UNIVERSITY OF SCIENCE TECHNOLOGY	大学(中国)	1447
23	UNIVERSITY SYSTEM OF GEORGIA	大学(米国)	1447
24	MICROSOFT	民間企業(米国)	1443
25	BEIJING INSTITUTE OF TECHNOLOGY	大学(中国)	1419
26	XIDIAN UNIVERSITY	大学(中国)	1327
27	PEKING UNIVERSITY	大学(中国)	1310
28	SOUTHEAST UNIVERSITY CHINA	大学(中国)	1309
29	UNIVERSITY OF TECHNOLOGY SYDNEY	大学(オーストラリア)	1287
30	PENNSYLVANIA COMMONWEALTH SYSTEM OF HIGHER EDUCATION PCSHE	大学(米国)	1285
31	UNIVERSITY OF NORTH CAROLINA	大学(米国)	1281
32	HELMHOLTZ ASSOCIATION	民間研究機関(独国)	1277
33	NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY CHINA	大学(中国)	1253
34	XI AN JIAOTONG UNIVERSITY	大学(中国)	1240
35	TECHNICAL UNIVERSITY OF MUNICH	大学(独国)	1237
36	STANFORD UNIVERSITY	大学(米国)	1220
37	NORTHEASTERN UNIVERSITY CHINA	大学(中国)	1208
38	BEIJING UNIVERSITY OF POSTS TELECOMMUNICATIONS	大学(中国)	1194
39	HARVARD UNIVERSITY	大学(米国)	1192
40	UNIVERSITY OF TOKYO	大学(日本)	1184
41	IMPERIAL COLLEGE LONDON	大学(英国)	1179
42	UNIVERSITY OF ILLINOIS SYSTEM	大学(米国)	1167
43	UNIVERSITY OF ELECTRONIC SCIENCE TECHNOLOGY OF CHINA	大学(中国)	1150
44	CONSIGLIO NAZIONALE DELLE RICERCHE CNR	国立研究機関(イタリア)	1149
45	ETH ZURICH	大学(スイス)	1144
46	NORTHWESTERN POLYTECHNICAL UNIVERSITY	大学(中国)	1135
47	SOUTH CHINA UNIVERSITY OF TECHNOLOGY	大学(中国)	1109
48	UNIVERSITY OF SCIENCE TECHNOLOGY OF CHINA	大学(中国)	1099
49	UNIVERSITY SYSTEM OF MARYLAND	大学(米国)	1065
50	CITY UNIVERSITY OF HONG KONG	大学(中国)	1064

出典: Web of Science Core Collectionでの検索結果をもとにNEDO技術戦略センターが作成(2018)

上位50機関中、論文数が多い機関の国籍順は、中国(22)、米国(13)、仏国(3)、英国(2)、独国(2)、シンガポール(2)、日本(1)、インド(1)、オーストラリア(1)、イタリア(1)、スイス(1)、イラン(1)であった。

次に、日本機関の順位を示す(表2-10-18)。順位欄には、世界全体における順位を示した。

■表2-10-18 人工知能技術論文数が多い研究機関一覧(国内機関)(2013~2018年)

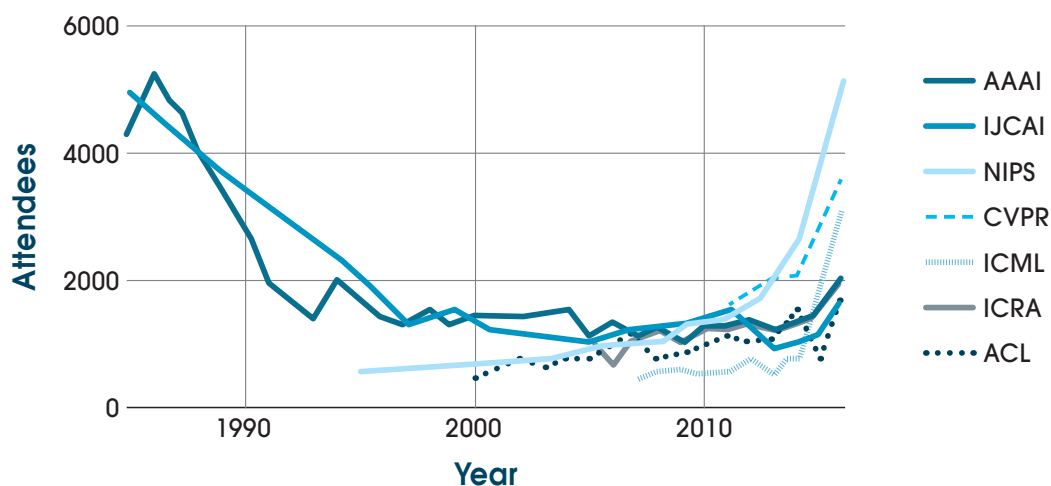
順位	機関	論文数
40	東京大学	1184
115	大阪大学	703
168	早稲田大学	553
228	京都大学	471
236	東京工業大学	464
241	名古屋大学	453
249	情報・システム研究機構	443
276	筑波大学	415
300	国立情報学研究所	387
309	九州大学	382
341	電気通信大学	352
348	産業技術総合研究所	347
351	東北大学	345
373	慶應義塾大学	328
387	奈良先端科学技術大学院大学	312
396	立命館大学	396
416	九州工業大学	297
443	北陸先端科学技術大学院大学	282
471	大阪府立大学	267
498	北海道大学	256
528	理化学研究所	245
608	神戸大学	216
655	広島大学	201
682	名古屋工業大学	193
712	首都大学東京	185
719	科学技術振興機構	183
745	東京理科大学	178
750	豊橋技術科学大学	177
770	日本電信電話(NTT)	173
779	本田技研工業	172
790	東京農工大学	169
823	日本電気(NEC)	162
826	情報通信研究機構	161
833	法政大学	159
838	会津大学	158
907	福岡工業大学	145
945	岩手県立大学	139
964	山口大学	135
964	富士通	135

出典: Web of Science Core Collectionでの検索結果をもとにNEDO技術戦略センターが作成(2018)

#### ④国際会議からみた技術動向

AI関連主要国際会議について、出席者の推移を図2-10-12に示す。AAAI、IJCAIは、記号推論などの伝統的なテーマを含む人工知能技術全般、NIPSはニューラルネットワーク技術、CVPRは画像認識、ICMLは機械学習を主要なテーマとして扱う学会である<sup>\*155</sup>。これらの会議は機械学習及び深層学習に関連した技術を主要なテーマとしている。国際会議ごとの出席者数の推移から、記号論理から機械学習及び深層学習へのシフトが起きたことが推察される。

■ 図2-10-12 AI主要国際会議の参会人数の推移



出典: AI Index2017 Annual Reports, Stanford Univ. (2017)

主要な国際会議である、AAAI、IJCAI、NIPSの最近の状況を以下に紹介する。

#### ・AAAI (Association for Advancement of Artificial Intelligence)

米国の人工知能技術に関する学会であり、IJCAI、NIPSと並んで人工知能分野のトップカンファレンスの一つに位置づけられている。2018年大会(2018年2月2日~7日米国ニューオーリンズにて開催)では、採択数は全体で938件であった(表2-10-19)。次に、国別投稿数及び採択数を示す(表2-10-20)。1位は米国250件(投稿数934件)、2位は中国240件(投稿数1,248件でトップ)であり、米国・中国で半数近くを占める。日本は30件であった。

■ 表2-10-19 投稿数及び採択率の推移

年	採択数	採択率(%)	採択数増加率 前年比 (%)
2014	398	28.3	-
2015	539	27.0	41.6
2016	548	25.5	7.6
2017	539	24.8	19.9
2018	938	24.6	48.1

出典: 公開情報をもとにNEDO技術戦略センターが作成(2018)

\*155 AAAI: Association for the Advancement of Artificial Intelligence, IJCAI: International Joint Conference on Artificial Intelligence, NIPS: Conference and Workshop on Neural Information Processing Systems, CVPR: Conference on Computer Vision and Pattern Recognition, ICML: International Conference on Machine Learning, ICRA: IEEE International Conference on Robotics and Automation, ACL: Association for Computational Linguistics



■表2-10-20 国別投稿数及び採択数

順位	国/地域	2018年採択数	2018年投稿数	2017年投稿数(参考)
1	米国	250	934	770
2	中国	240	1,248	785
3	シンガポール	40	109	94
4	香港	34	92	61
5	豪州	32	124	104
6	英国	32	128	101
7	日本	30	99	80
8	独国	(公表なし)	68	60
9	インド	(公表なし)	76	58
10	カナダ	(公表なし)	70	70

出典:公開情報をもとにNEDO技術戦略センターが作成(2018)

分野別には、機械学習が1/4を占め、最多数であるが、従来の人工知能分野も広く採択されている(表2-10-21)。

■表2-10-21 分野別論文数

順位	分野	採択数	投稿数
1	機械学習	261	1,164
2	ビジョン	128	435
3	自然言語と機械学習	73	262
4	人工知能とWeb	57	236
5	機械学習の応用	46	227
6	ゲーム理論と経済パラダイム	45	138
7	自然言語とテキストマイニング	44	182
8	知識表現	41	163
9	アプリケーション	40	207
10	マルチエージェント	25	98
11	自然言語と知識表現	22	87

出典:学会事務局による発表

#### ・IJCAI(International Joint Conference on Artificial Intelligence)

人工知能分野でのトップの学術会議であり、AAAIと並び最難関とされている。2018年の開催地はスウェーデンのストックホルムであった(2018年7月13日~19日)。採択数、分野別論文数を以下に示す(表2-10-22、表2-10-23)。

■表2-10-22 国別採択数

国/地域	採択数
中国	325
EU	129
米国	122
シンガポール	26
オーストラリア	23
日本	17
イスラエル	13
香港	12
インド	10
カナダ	10

出典:学会事務局が発表した採択リストをもとにNEDOが作成

■表2-10-23 分野別論文数

分野	採択数	投稿数
機械学習	356	1,808
機械学習の応用	131	613
コンピュータービジョン	131	613
自然言語処理	102	480
マルチエージェントシステム	100	498
知識表現	81	412
計画とスケジューリング	44	235
探索とゲーム	42	190
制約と充足	41	162
人工知能の不確実性	41	188
人間と人工知能	38	260
ロボティクス	25	136

出典:学会事務局が発表した採択リストをもとにNEDOが作成

### ・NIPS (Conference on Neural Information Processing System)

当初は神経科学を計算科学の側面から研究するという方向性の発表もあったものの、現在はほとんどが機械学習の発表に占められており、機械学習のトップレベル会議とされている。

2017年大会(2017年12月4日～9日米国ロングビーチにて開催)の採択数を以下に示す(表2-10-24)。

■表2-10-24 国別採択数

国/地域	採択数
米国	406
中国	163
カナダ	22
日本	14
欧州	163
アジア(日本除く)	26
その他	23

出典:学会事務局が発表した採択リストをもとにNEDOが作成

当該大会では、全体的に以下の傾向が見られた。

- ・ Generative Adversarial Net (GAN)に関する発表の増加
- ・ ベイズ推論では、Variational Inference (VI、変分ベイズ法)の発表が増加
- ・ 「メタラーニング」がキーワードになりつつある(GAN、深層強化学習もこれに含まれる)
- ・ AIチップの開発が進展

国際トップカンファレンスにおいて、日本からの発表が数の上ではそれほど多くないが、AAAI2018の採択率は決して低くない。研究者の層を厚くし、また外国研究グループのネットワーク形成などの取組みの充実が望まれる。

## 2.10.2 グランドチャレンジ

### (1) 総論

現在までの人工知能(Artificial Intelligence; 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、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生み出してきた。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ー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(米国)のDeep Blueが、当時の世界チャンピオンであるGarry Kasparov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。

これらのチャレンジが、ゲームの世界を対象にしていたのに対し、物理世界での課題をテーマとするチャレンジも登場している。

我が国の研究者が中心となり1990年代半ばに始まったRoboCup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(Fédération Internationale de Football Association)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カ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。これは、サッカーというテーマを設定しているがゆえに、複数の自律型ロボットの実時間での協調動作という新たな課題を提示した。この問題設定は、人工知能とロボットの複合領域での研究と実用化を加速した。これは、RoboCup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ことでも見て取れる[1]。また、米国国防高等研究局(DARPA)も自動走行車や災害救助ロボットの領域でグランドチャレンジを設定して、数回の競技会を開催したことがある。

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。RoboCupを例にとると、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3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。

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生み出され、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]。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。

## (2) 完全情報問題をテーマとしたグランドチャレンジ

これまで、人工知能のグランドチャレンジは、チェス、将棋、囲碁など、ゲームを対象に設定されてきたことが多い。ゲームは、すべての情報が眼前にあるという意味で、完全情報問題であり、さらに、ルールが明確であり、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の例題としてチェス(の世界チャンピオンに勝つコンピューターを開発すること)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。John McCarthy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。

当初は、チェスの解空間は、あまりに膨大であり、力ずくの探索は現実的ではなく、人間の持つ知識をヒューリスティックスとして組み込む必要があるとの議論がされていた。しかし、その後の展開は、これとは全く逆であった。チェスは、探索アルゴリズムの高度化、大規模並列計算機による探索の地平の拡大、過去の棋譜を用いた盤面評価の学習の3要因が組み合わさることで急速に強くなっていった。

Claude ShannonとAlan Turingはチェスの探索にゲーム理論でJohn von Neumannらが開発したミニマックス法<sup>\*156</sup>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、

---

※156 何らかの評価関数にもとづき、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。

- 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方が強くなるので、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ことになる。

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すべ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が同じで、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ものをアルゴリズムとしてまとめたのがDonald Knuthである。Knuth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よいことを明らかにした。

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よいことになる。チェスのプログラムで経験的にそのことが分かり、それが後に反復深化(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<sup>\*157</sup>、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)ルール上指せるすべ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ー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。ま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、コンピューターは(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。コンピューターはこれをも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駒のすべ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ー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Kasparovに勝利した。6回戦勝負で、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Kasparov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Kasparovの実力はDeep Blueに勝っていたと思われる。

とは言うものの、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

---

\*157 キーと値をペアで管理するデータ構造。

ーター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ー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。

## 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 (すでに人間よりもコンピューター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ははるかに大きく、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ー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<sup>\*158</sup>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すべて使った。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年前後にはすで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か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 (現在、電気通信大学) のBONANZA (ボナンザ) によってコンピューター将棋は革命的な進歩を果たした。保木の工夫は、

- 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ー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すべてが取り入れており、「ボナンザメソッド」と呼ばれている。保木はボナンザ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ー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ーとプロ棋士が対戦するようになった。2013年、2014年と電王

---

\*158 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。

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ーが圧勝した。この時点ですでにコンピューター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ー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年に山本一成が開発したPONANZAが佐藤天彦名人に圧勝している。将棋においても、チェスと同様に、高度な探索アルゴリズム、膨大な計算量、盤面評価の学習が重要な要因となっている。

チェスにおいても将棋においても、「確かに人間を凌駕するシステムは開発できたが、人間のように考えているわけではない」という議論がなされた。しかし、これらのチャレンジは、チェスや将棋というゲームにおいて、人間を凌駕するシステムができるかというチャレンジであって、人間の思考過程を模倣するチャレンジではないのである。また、人間のトッププロは、すべての過去の棋譜を覚えており、圧倒的に先読みが優れ、盤面評価が正確である。これらは、人工知能システムと共通の要素である。残る部分は、いわゆる「ひらめき」の部分であり、今後は、「ひらめき」と言われている部分が何かに焦点が集まってくるであろう。しかし、すでに人工知能システムが、予想外の奇手を打ち、それが極めて有効であるという事例も生まれている。そうすると、「ひらめき」は、想定外の解空間への探索によって実現されるという考え方も成り立つであろう。

### 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。最初にコンピューター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von Neumann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Remi Coulom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いちばん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入っていない。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6段程度の実力に達した<sup>\*159</sup>。

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

---

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

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月にGoogle（米国）のAlphaGo（碁）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。AlphaGoは、

- 1) 深層学習
- 2) モンテカルロ木探索
- 3) 強化学習

という3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深層学習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さらに強くした。これまでコンピューター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Lee Sedol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も深層学習を取り入れてトッププロ棋士と良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。

AlphaGoで採用された技術的手法は、チェスや将棋のようにその場で、探索を行うものではなかった。極めて多くの盤面に対して、ベストな打ち手を事前に学習している方式であった。

AlphaGoの場合、AlphaGo対AlphaGoという対局を膨大な回数行い、今までの棋譜に現れていない盤面からの展開とその評価を事前に学習している。実際の人間との対局の段階では、人間のトッププロよりAlphaGoのほうが、はるかに多くの経験を積んでいるのである。

その後開発された、AlphaGo Zeroは、棋譜を全く使わず、白紙の状態からランダムに手を打ちながら学習を進める方式をとり、数日で、AlphaGoを凌駕するまでに至った。その過程で、「劫」など囲碁における基本規則を獲得している。AlphaGo Zeroの意味することは重要である。棋譜を使わないで、棋譜を使ったシステムを凌駕したということは、いわゆるビッグデータの有効性の再検討が必要となる。棋譜は、今まで人間が行ってきた囲碁の解法集ともいえる。しかし、囲碁というゲームを考えたとき、その解法集が解空間の一部に偏っている可能性も否定できない。その場合、その解法集を利用せずに、ランダムに解空間全体を探索したほうが、より良い解を発見する可能性があるということの意味する。実際に、AlphaGo Zeroが、AlphaGoを凌駕したということは、棋譜は、人間の思考のパターンに沿った囲碁の解法集であり、囲碁というゲームの最も強力な解法集ではなかったとも言える。これは、今後、我々は、学習データの収集と生成、システム全体の設計原理を考えるうえで極めて重要なことである。

### (3) RoboCup

RoboCup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。RoboCupは、浅田稔、野田五十樹、北野宏明、松原仁ら、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 (RoboCup Rescue) や教育 (RoboCup Junior) に関する活動が加えられた。では、家庭用ロボット (RoboCup@Home) や物流など産業用途ロボット (RoboCup Industrial) をタスクとしたリーグが増えてきている (表 2-10-25)。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。45カ国から、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 (RoboCup2018では35カ国、約4,000人が参加)。

■表2-10-25 RoboCupのリーグ構成(2018)※160

RoboCupのリーグ種別	概要
RoboCup Soccer	サッカー
Humanoid League	人型ロボット
Middle Size League	車輪中型
Small Size League	車輪小型
Simulation League (2D / 3D)	シミュレーション
Standard Platform League	標準型ロボット
RoboCup Rescue	災害救助
Robot League	災害救助ロボット
Simulation League	シミュレーション
RoboCup Junior	教育
Rescure League	災害救助
Soccer League	サッカー
Onstage League	ステージパフォーマンス
RoboCup@Home	家庭内向けロボット
RoboCup Industrial	産業用途ロボット
RoboCup Logistics	物流ロボット
RoboCup@Work	オフィス環境向け

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、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。

RoboCup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世紀中ごろに、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(不完全情報<sup>※161</sup>など)を抽出し、そのうえで、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、誰にでもひと言

※160 「RoboCup 2018Webサイト <<http://www.robocup2018.org/>>

※161 意思決定時において、必要な情報が不完全であること。

で理解してもらうことが可能で、さらに研究者自身が熱くなれるテーマとして選ばれた。

これから必要となる技術は、チェスや将棋のように、すべ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。不確実な情報をもと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であろう、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、世界中で受け入れられるテーマとして、サッカーが選ばれている。

RoboCup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。RoboCup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、7億7,000万ドル(約800億円)という大きな評価額でAmazon(米国)に買収され、さらにRoboCup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、ソフトバンクから1億ドルの出資を受けるといったことが起きたのである。

KIVA Systemsは、コーネル大学(Cornell University、米国)のチームを率いたRaffaello D'Andrea(現在、チューリッヒ工科大学<Eidgenössische Technische Hochschule、スイス>)らが、RoboCup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、RoboCup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。D'Andrea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。

RoboCup Rescueは、RoboCup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みも必要であるとの認識から始まっている。

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うえに、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によって、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カ月後、9.11のテロが発生し、米国から参加していた南フロリダ大学(University of South Florida)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であったが、テロ発生のお知らせに、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。RoboCup Rescueは、その構想段階から米国の連邦緊急事態管理局(Federal Emergency Management Agency; 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RoboCupに関心を寄せていたアメリカ国防高等研究計画局(DARPA)の強い興味を引きつけ、RoboCup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。また、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、RoboCup Rescue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。RoboCup Rescue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、さら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。

RoboCup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さらに広範に援用した、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。

#### (4) DARPAにおけるグランドチャレンジ

DARPAとは、米国防総省の研究開発機関であり、米国における人工知能研究を支えてきた存在でもある。DARPAは、2000年ごろから、グランドチャレンジと呼ぶ競技会の実施を構想し始めた。その後、2004年ごろから、DARPA Grand Challenge(2004年、2005年)、DARPA Urban Challenge(2007年)、DRC(DARPA Robotics Challenge)が実施された。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。

その後、災害救助をタスクとした、DRCが開催された。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、プログラマネージャーであるGill Pratt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4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すべてを開発する

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すべてを開発する

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、コンピューター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、優勝を取めた。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併せて、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、タスクの数と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ドアを開けて室内に入る、④バルブを回す、⑤工具を持ち、壁に穴を開ける、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、転倒防止策なし)

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ー間の通信路は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、さらに情報はロボットからオペレーターへの一方向でしか送れない通信路である。オペレーター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の結果は表2-10-26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2-10-26 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 RESCUE	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すべてが1度は競技中に転倒したことも合わせて考えると、二足歩行はさら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。

DARPA Grand Challengeは、自動走行や災害救助ロボットなどをテーマとして設定し、注目を集めているが、グランドチャレンジと呼称するには、継続性に乏しいという問題がある。DRCも現在は、開催されておらず、その目標へと到達するようなプログラムにはなっていない。グランドチャレンジをテーマとした、ショーケース的競技会が行われる程度である。これは、DARPAが主催するということや、プログラム・マネージャー制度のため、プログラム・マネージャーが変わると継続される保証がないという現実がある。その点、RoboCupは、グランドチャレンジのための法人が設立され、国際的な運営委員会が設置されるなど、継続の仕組みができあがっている。グランド・

チャレンジでは、その継続の仕組みをしっかりと構想することも重要である。

### (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、RoboCupの提唱者の一人でもあるソニーコンピューターサイエンス研究所 (Sony Computer Science Laboratories; Sony CSL、日本) の北野宏明は、「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 [2]。特に、医学生理学賞をターゲットとしている。

さらに、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、すで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つであるなど、本当の意味で大きな科学的発見に結びつ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ーが実現している。同時に、1990年代中ごろ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にある」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では、大規模データ、大規模計算、さらには、機械学習という3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<sup>\*162</sup>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思われる。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であろうと思われる。

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.

※162 北野宏明ら、システムバイオロジー研究機構の開発チーム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。

## 2.11 ▷ 今後の展望

本章では、近年のAIの技術及び研究開発環境の動向を整理した。ABCIなどの開発基盤、人材育成、国際標準化などの開発環境関連については、前回の『AI白書2017』発刊からの1年余りの間、速度を緩めることなく進展していることが感じられる。AI技術に関しては、新しいAIのコンセプトや手法の登場といった大きな変化は少ないものの、人間のように流ちょうに話す技術や判断の理由を人間に説明する技術、身体性とAIの関係を軸としたロボティクス技術など、実利用に近い領域での技術や手法については着実な進展が感じられる。

このような現状を鑑みると、本白書の第5章でも取り上げているように、開発環境の充実が進展する下で、技術の社会実装をいかに進めるかが重要な課題と考えられる。今後のAI白書においても、技術及び研究開発環境の動向を把握し、次章の利用動向と照らし合わせることによって、社会実装状況の確認を行っていくことが必要と想定される。

# MEMO



# AIによるクリエイティブの可能性

株式会社ドワンゴ  
川上量生

芸術のように人間が生み出す創作活動は、これまでコンピューターに置き換えることが最も難しいものの代表例とされてきた。

そもそも人間がどのように創作活動を行っているか自体、断片的なノウハウとしてクリエイターたちによって語り継がれているものはあっても、基本はそれぞれのクリエイターによって異なる“べき”ものとされていて、クリエイティブの統一基本法則のようなものは一般には共有されていない。

音楽理論や脚本術など、比較的理論的に整理されているように見えるジャンルはあっても、それは基礎的な技術を教えるだけであって、人間が人間を感動させるクリエイティブの最も貴ぶべき神髄は、人間のもつ「ひらめき」や「直感」、何か神秘的な「才能」なるもので説明されることが多く、およそ、科学的な説明を拒んでいるようにもみえる。

しかし、その固定観念に疑問を投げかけるような結果が、近年のディープラーニングの研究によりもたらされている。人間のもつ「ひらめき」や「直感」なるものをコンピューターで再現することは、想像以上に簡単なメカニズムである可能性がディープラーニングによって示されているといってもいい。

人間は頭の中から、現実にはない空想の世界を作り出すことができる。ある種のひらめきによって、そういったことを行う方法論の一つとして、なにかから「連想」をするという手法がある。あるものの特徴から、別のものとの共通点を見つけ出して結び付けるというやり方だ。こういったことは人間だけではなく、AIでもできることが分かった。

Google(米国)が2015年に発表した「Deep Dream」[1]が最初の例になるだろう。AIは与えられた画像の中からパターンを見つけ出し、そのパターンに似た特徴をもつ別の絵を描く。山や木から不思議な塔が出現し、葉っぱからは鳥の姿が浮かび上がる(図1)。人間でいうと、天井板の木目を見つめていると、人間の顔などが見えてくる、とかいうことがあるが、同じことがコンピューターにも再現できたというわけだ。いわば機械による「連想」が実現した。さらに人間の一部は絵描きとして、そういう連想を映像に落とし込む能力をもっているわけだが、それもAIができてしまった。人間の全員ができるわけではない創作活動の一つがAIでもできるようになったのだ。

「Deep Dream」を見ると、どこか非人間的な狂気を感じる。人間ではないものが作ったのだから、当たり前だということもできるかもしれない。しかし、おそろしいのはそういった作品から狂気を感じさせるという現象は、これまで人間の世界ではある種の超越した才能をもつ作家だけができる“天才性”の証明であるとも考えられてきたことだ。

クリエイターのもつ神秘性の核心の部分に、いきなり「Deep Dream」は切り込んできたといえる。

■図1 GoogleのDeep Dream

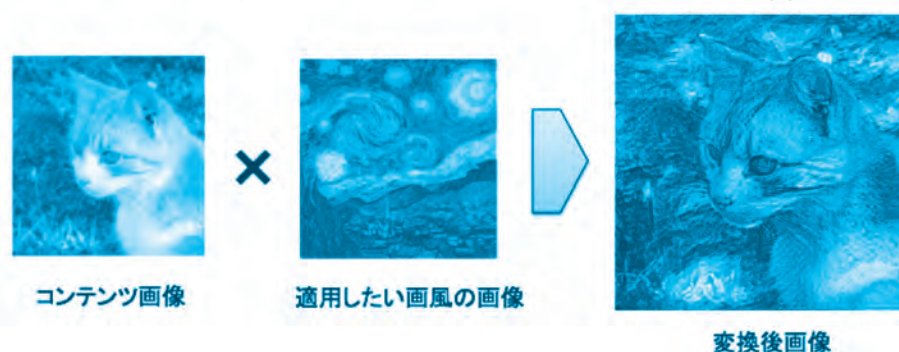


出典: Google AI Blog  
 「Inceptionism : Going Deeper into Neural Networks.」(1)(クリエイティブコモンズライセンス、表示4.0国際 <https://creativecommons.org/licenses/by/4.0/>)

もう一つの人間の創作能力の例として「アレンジ」がある。音楽では、ある楽曲をテクノ調にしたり、ヒップホップ風にしたりといったことは、ディープラーニング以前にも、ある程度機械的に行う手法が存在した。しかし、ある画家のタッチに似せて「○○風に描く」という行為は、訓練された人間の画家には簡単にできても、コンピューターでどうやって実現すればいいのか、まったくめどがたっていなかった。

それがディープラーニングを用いて、簡単に実現した例が現れた(図2)。チュービンゲン大学(ドイツ)の研究発表[2]をもとにした、Preferred Networks(日本)の事例[3]によると、猫の写真と絵画の写真をAIに渡して、その絵画風に猫の絵を描く、ということが実現できた。ゴッホのような人間の画家風に描いた猫の絵だけではなく、新聞記事を画風と見立てて、まるで新聞記事のように見える猫の絵なども同時に公開された。こうなると、ちょっと「……風」に描くという技術は、人間よりもすでにAIのほうが得意な分野に思える。

■図2 テュービンゲン大学の研究をもとにしたPreferred Networks による画風転写の事例



出典: Preferred Networks Research ブログ「画風を表現するアルゴリズム」[3]

もう一つNVIDIA(米国)の2017年末の研究発表を紹介する[4]。人間はレイアウトを指定すると、それを埋めるように絵を描いて、現実にはない映像を作り出すことができる。手描きアニメーション制作の現場では日常的に行われている光景だ。

こういったこともAIにやらせることができた。人間が領域を色分けして指定すると、あとはそれに従ってAIが自動的に絵を作成することもできるようになった(図3)。しかも、人間並みどころか、写真と見分けがつかないレベルのものをAIが描くということが現実となった。

■ 図3 NVIDIAによる、領域指定画像(左)とそれにもとづく自動生成写真(右)



出典:NVIDIA "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs" [4]

このような事例を見ると、クリエイターにできてコンピューターにできないことというのは、クリエイティブを成り立たせている構成要素としては、急速になくなりつつあるように見える。しかも、一旦コンピューターができるようになったことだけで見ると、細部にわたる緻密な創作は、むしろ人間よりも得意なのだ。

こういった現実には、人間のクリエイティブの方向性そのものに影響を与える可能性がある。

映像作家の庵野秀明監督に、こういったAIを使った映像技術について紹介したときに、興味深いコメントをしてくれた。それはつまり「映像技術の進歩は、新たな演出方法の発明とは別に科学技術の進歩の恩恵によるところが大きい。無声から音が付いたり、白黒から色が付いたり、平面から立体的になったり、アナログからデジタルになったりと、常に最先端技術に合わせて映像表現も変化している。その流れに沿うと、近々AIの技術を取り入れた映像が主体になっていくかもしれないですね。」

さらに続けて「特にアニメーションは情報のコントロールが特徴なので、AIの恩恵を受けやすい気がします。AIを利用した映像処理のコストダウン等もありますが、僕はAIを使ったプロットやシナリオ、果ては作品制作そのものも可能になるのでは、と感じています。個々の映像作家は元より、観客自体も自身の嗜好性や好きな思考、快樂原則はほぼ決まっているので、その個人向けパラメーターや大衆娯乐的なビッグデータを使った、個別対応なパーソナル映像作品がAIにより作られていくのではないのかなと。その人が観たい理想的な映像を常に学習して情報を更新しつつ作り上げていく、SFの世界みたいなことが起こり得るのではないのかなと。そうすると人間はそのための素材作りか、そのシステムに抗って今と同じシステムで作品を作ってるんじゃないですかね。」

庵野監督のような時代を代表する一流クリエイターが、映像作家自身だけでなく観客自体についても嗜好性や思考、快樂原則はほぼ決まっていると断じているのは、非常に興味深い。クリエイティブの重要な部分として、ある種のパラメーターの調整のようなものが存在していて、それはビッグデータやパーソナライズを駆使できるAIに、将来は置き換わる可能性があるということだ。

クリエイター自身を囲碁将棋AIのような学習機械であると仮定すると、なにかを極めている達人ほど、長年の試行錯誤の結果、クリエイティブについてのシンプルで整理された評価関数を自分の中で持っているとは想像することはできないだろうか。

将棋界では、大山康晴名人が「コンピューターに将棋を指させてはいけない」と言ったという話が有名だ。人間がコンピューターに勝てるわけがないというのだ。おそらく大山名人のような達人は、将棋の評価関数が自分の中で明確に定義できていたのではないだろうか。そして、そういった関数が分かっただけで、人間よりも高速かつ大量に計算できるコンピューターには、人間が負けると思っ

ていたのではないか。

将棋や囲碁と似たような構造、あるいはその延長線上で、AIがコンテンツを作る日が来るのではないかと私は考えている。しかし、人間の心を感動させるコンテンツがAIなんかに本当に作れるのだろうか。

突然現れて大ヒット作品を作り出す若い天才作家につけられる形容詞として「時代を捉えた若い感性」というようなものがある。また、そういった若い才能が、年を取るにしたがって、時代の変化についていけずに色あせていくという現象もよくみられる。

こういった現象は、たまたまその若い作家が、同時代の人々がどういう作品に反応するかについての最大公約数的な感性を、自分自身で持っていたというように解釈すべきだろう。ビッグデータがネット経由で収集できる時代には、時代を読んだヒット作品を定量的に予測することも、AIのほうが得意である可能性が高い。

そう考えると、商業作品の創作活動においてAIの役割が相当に大きくなっていくだろうことは容易に想像できる。

また、完全にAIに創作活動を任せないのだとしても、AI、特にディープラーニングは、人間がどのように発想すればいいのかのヒントを教えてくれるようになるだろう。また、ディープラーニングで分かった成果をクリエイティブに活かすということも可能になるはずだ。

DeepZenGoの開発で協力を頂いた囲碁棋士の藤澤一就八段は、コンピューター囲碁ソフトのやり方を人間の棋士も学ぶべきだと私に語ってくれた。コンピューターは人間よりも遙かに多くの手筋を読むので、新しい局面でどういう風に打てばいいのかという研究に使っている棋士は現在でも多いのだそうだ。中国や韓国の棋士の強さはある意味、コンピューターのシミュレーション能力に似ているので、後半の正確度が高く、現状、日本の棋士はなかなか勝てない。だが、読みの力は年齢とともに衰えるので、30代以上で活躍するのは中国や韓国の棋士の場合には非常に難しく、シニア棋士だけで国際棋戦をやれば、まだ、日本のほうが強いのではないか。年を取っても日本の棋士の強さが落ちにくいのは、大局観に優れているからだ。囲碁ソフトは大局観をさらに鍛える道具として使えるはず、というのが藤澤八段の主張だ。

具体的には囲碁ソフトの評価関数の値が大局観を養うのに使える、という。ある局面が有利か不利か、そういうことを直感的に判断できるような感覚は、評価関数の値を参考にすることで大局観を養い、確認することができる。詰め碁のように読みの力を強める練習方法はたくさん存在しているが、大局観を養う練習方法はそれほどない。それが囲碁ソフトを使って補強できるという。すでに弟子たちの練習に取り入れているそうだ。

AIが人間をあらゆる意味で超える日が、いつか来るのだろう。しかし、人間がAIに学び、さらに能力を向上させるという時代も、相当に長いのではないかと、私は思っている。

#### ◆参考文献

- [ 1 ] "Inceptionism: Going Deeper into Neural Networks." Google AI Blog<<https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>>
- [ 2 ] "A Neural Algorithm of Artistic Style." arXiv.org<<https://arxiv.org/abs/1508.06576>>
- [ 3 ] "画風を変換するアルゴリズム" Preferred Networks Researchブログ <<https://research.preferred.jp/2015/09/chainer-gogh/>>
- [ 4 ] "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs." arXiv.org <<https://arxiv.org/abs/1711.11585>>

# 機械学習工学

株式会社Preferred Networks

丸山 宏

## プログラミングパラダイムとしての統計的機械学習

人工知能の研究開発で生まれた技術のうち、いくつかのものは日々の情報技術の中に取り込まれている。例えば記号処理のために開発されたガーベージコレクションなど動的メモリ管理、自然言語処理のために作られた構文解析アルゴリズム、ゲームなど解空間の中で解を求めるための探索アルゴリズム、それに知識表現のために盛んに研究された概念階層などである。今では、これらの技術を「人工知能」と呼ぶ人は少ないだろう。同様に、深層学習に代表される統計的機械学習も、今後の一般のプログラミングに取り込まれていくものと予想される。

ここでは統計的機械学習を、人工知能の文脈ではなく、新しいプログラミングパラダイムとして情報技術の中に取り込まれていく技術として捉え、その技術を有効に使うための工学的知識体系（本稿では機械学習工学 [1] と呼ぶ）を構築する動きについて述べる。

統計的機械学習をプログラミングパラダイムとして捉えるとは、関数  $y=f(x)$  をプログラミングする際に、 $f$  の計算手順を書き下すのではなく、 $f$  が行う計算の入出力の例示  $\{ \langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle, \dots, \langle x_n, y_n \rangle \}$  を与えることによって、 $f$  を帰納的に定義することを指す。この入出力の例示を「訓練データセット」と呼ぶ。 $f$  の計算手順を書き下す伝統的プログラミングとは異なり、統計的機械学習による  $f$  の実装は、訓練データセットからほぼ自動的に作られる。

統計的機械学習の一つである深層学習では、 $f$  の実装は入力値が伝播していく深層ニューラルネットの形で表現される。与えられた任意の関数、任意の精度に対して、十分な数の中間ノードを持つニューラルネットが存在して、その関数をその精度で近似できる [2]。従って深層ニューラルネットは計算モデルとして擬似的にチューリング完全（すべての計算可能関数を表現可能）と言える。「擬似的」というのは、チューリングマシンと異なり、結果が近似でしか得られないからである。深層ニューラルネットは任意の関数を近似できるが、それが現実的な量の訓練データと計算時間で訓練できるかどうかは別の問題である（普通のチューリングマシンにおいて、アルゴリズムは書いても現実的な計算時間で終わらないものが存在することに似ている）。

統計的機械学習による帰納的なプログラミングは、チューリングマシン抽象にもとづく従来のプログラミングモデルとは根本的に異なる技術である。この新しいプログラミングモデルを効果的にかつ安全に使いこなすにはどうしたらよいか、まだ試行錯誤している段階である。特に、深層学習による画像認識技術が自動運転車などに使われ始め、深層学習システムの構築ができる人材の不足や、システムの品質や安全性に対する懸念が議論されている。

この状況は、デジタル計算機が普及し始めた1960年代の状況によく似ている。当時IBMがSystem 360という汎用機を商用化し、そのソフトウェアが大量に作られたが、プロジェクト予算

や納期の超過、低い品質、保守の困難性など様々な問題が発生し「ソフトウェア危機」と呼ばれた。

この「ソフトウェア危機」はソフトウェア工学の発展により、その後解消していった。同様に我々は今、統計的機械学習を安全かつ効果的に使いこなすための知識体系すなわち機械学習工学を必要としている。我が国においては、2018年4月に日本ソフトウェア科学会の研究会の一つとして「機械学習工学研究会（主査、石川冬樹国立情報学研究所教授）」を設置し、活動を始めた。機械学習工学研究会は、ソフトウェア工学と機械学習の研究者・実務家の意見交換の場を提供し、機械学習工学の様々な経験・知見を知識として体系化する支援をしている。

## 機械学習工学の課題

機械学習工学はまだ始まったばかりの議論なので十分に体系化はできていない。ただし、シンポジウムやワークショップで繰り返し取り上げられるトピックがあるので、それらを3点紹介する。

### (1) 機械学習応用システムの開発・運用プロセス

統計的機械学習を用いたシステム（機械学習応用システムと呼ぶ）の開発には、訓練データセットからの訓練が必要だが、訓練後のモデルがどのような精度になるかは、やってみないと分からない。精度は訓練データセットの量、質、データのばらつきなどに大きく影響されるからである（特に、訓練データセットあるいはその一部として、すでに得られているデータを与えられる場合には、そのデータに外れ値や欠測値がないか、バイアスがないかなどを十分に吟味する必要がある）。このため、機械学習応用システムの開発は試行錯誤を含む探索的なプロセスとなる。

このような探索的開発には、事前に要求定義を決めて開発を行うウォーターフォール型の開発プロセスはそぐわない。必然的に、短いサイクルで目標を見直すアジャイル型の開発プロセスを使うことになる。

売上予測モデルのように、時とともに季節要因や人口動態によってモデルが前提としていた条件が変わっていく場合も多い（これを「コンセプトドリフト」と呼ぶ）。コンセプトドリフトが考えられる応用では、開発が終わって運用に入った後も、訓練済みモデルは随時アップデートすることが求められる。従って、開発と運用を一体と考えるDevOps<sup>\*1</sup>の手法を取り入れることも必要である（MLOpsと呼ぶこともある）。

このように、機械学習応用システムでは今までのIT開発のやり方と変わってくるので、特に開発を外部に委託する場合には、現在の契約のプラクティスをそのまま利用するには問題がある。このため、経済産業省では「AI・データ契約ガイドライン」[3]を作成し、機械学習応用システムの委託契約において注意すべき点を明確にしている。この中で、特に、データ提供者の権利と義務、及び訓練済みモデルの帰属についてはよく考える必要がある。

現在の政府調達では、契約時に納入物の仕様についてコミットする請負契約が基本であるため、機械学習応用システムを開発して政府に納入することは難しい。アジャイル型の準委任契約が広く認められるようになることを期待したい。

---

※1 ソフトウェア開発において、開発（Development）と運用（Operations）が協力し、ビジネス要求に対して、より柔軟に、スピーディに対応できるシステムを作り上げるためのプラクティスをDevOpsと呼んでいる。

## (2) 品質保証

一般に訓練データセットは可能な入力値の集合のごく一部をカバーしているのに過ぎない。訓練データセットに現れない入力点に対する出力については、機械学習の汎化性能に任されていて、どのような値であるべきかをコントロールすることはできない。

このため、機械学習応用システムでは入出力の関係の厳密な仕様を書き下すことが不可能であり、これが品質保証を難しくしている。

入出力関係の厳密な仕様がないために、例えばテストケースを作ろうとすると、「仕様上正しい出力」すなわちテストオラクルを与えることが困難となる。また、出力がある不変量を満たしているかを保証することが難しい。さらに、出力が変わる境界のテストケース（いわゆるコーナーケース）をどのように定義するかも課題である。

一方で、機械学習モジュールは品質については、通常その精度で評価するために、定量的な評価を行いやすい。訓練データセットを作る際に、データの一部を評価に使うために取り分けておいて（これをホールドセットと呼ぶ）、これは訓練には使わない。

このホールドセットを用いて訓練後の機械学習モジュールの精度を評価すれば、客観的な評価が得られる。通常のプログラミングではできあがったモジュールの客観的な品質評価が難しく、多くはその開発プロセスを間接的に評価することになる。この点では機械学習モジュールの評価はより客観的であると言える。

## (3) 開発・運用環境

機械学習応用システムの開発・運用環境はまだ整備が始まったばかりである。機械学習モジュールの開発については、「深層学習フレームワーク」と呼ばれるライブラリが整備されつつある。代表的なものはTensorFlow、Chainer、PyTorchなどであり、いずれもオープンソースで開発が進められている。

機械学習モジュールは、コードと訓練済みモデルからなる。訓練済みモデルは、訓練コード、訓練データセットと、訓練に使われたハイパーパラメーター、それに訓練の初期値・乱数初期値の組によって一義的に決定される。できあがった機械学習モジュールが期待する振る舞いをしない場合は、これらのどれに問題があるか切り分けなければならない。そのためには訓練済みモデルがどのような訓練コード、訓練データセット、ハイパーパラメーター、初期値によって作られたかを紐づける、追跡システムが開発環境として欠かせない。また、機械学習の訓練には非常に大きな計算パワーが必要であり、現在ではその計算はGPUによって提供されている。従って、多くのGPUからなる計算リソースをどのように効率的に使うか、計算資源の共有と管理も機械学習工学の大きなテーマとなる。

開発が終わって運用フェーズに入ると、訓練済みモデルは、デプロイ環境に展開されることになる。訓練に必要な高額なGPU環境に比べて、機械学習モジュールのデプロイ環境は、エッジデバイスであったり、クラウドでもGPUのない環境であったりする。このため、訓練済みモデルを多様な環境にデプロイし、またそれらの精度を継続的にモニタリングする環境が必要となる。

## 工学へ向けて

プログラミングパラダイムとしての統計的機械学習は新しい技術であり、それを安全に効果的に利用するための工学的知識体系がまだ整っていない。UC BerkleyのMichael Jordan教授は、この状況を土木工学が確立されていないころのビルや橋の建築と同様ではないか、と指摘している [4]。機械学習応用システムの開発・運用のベスト・プラクティスが工学として体系化されれば、この新しい技術も社会に受容されていくに違いない。

### ◆参考文献

- [ 1 ] 丸山宏, 機械学習工学に向けて, 日本ソフトウェア科学会第34回大会予稿集, 2017.
- [ 2 ] Cybenko, G., "Approximations by superpositions of sigmoidal functions," *Mathematics of Control, Signals, and Systems*, Vol. 2, No. 4(1989), pp. 303-314.
- [ 3 ] 経済産業省, AI・データの利用に関する契約ガイドライン, <http://www.meti.go.jp/press/2018/06/20180615001/20180615001-1.pdf>.
- [ 4 ] Michael Jordan, "Artificial Intelligence—The Revolution Hasn't Happened Yet", *Medium.com*, Apr. 19, 2017.



# 構成的計算神経科学

国立大学法人大阪大学

浅田 稔

今回のAIの第三次ブームが過去と異なる決定的なポイントは、深層学習に代表されるニューラルネットワークのアーキテクチャーが活用されている点である。これまでは、人工知能と自然知能の乖離が問題であることを指摘していたが[1]、図らずも、深層学習が埋めるどころか、まさに現在のAIをリードしていると言っても過言ではない。「2.5.4 認知発達ロボティクスにおける身体性と社会的相互作用、自己概念の発達」で紹介した認知発達ロボティクスの例は、神経科学の知見を利用しているが、神経科学への貢献は少なく一方通行的に見える。本来は双方向であるべきである。そのつなぎの役割の一端を担うのが計算神経科学<sup>\*1</sup>である。本コラムでは、幅広い神経科学の分野の中でも、情報処理の観点から人間の知能に迫る計算神経科学のアプローチを簡潔に紹介し、そのうえで、「構成的」と称する意味合いを説明し、構成的計算神経科学の例を通じて、人工知能、神経科学、ロボットをつなぐ学際的アプローチとしての意味合いを探る。

## 1 計算神経科学とは？

銅谷は、計算神経科学を以下のように説明している[2]。

「計算神経科学は、脳が対処している感覚、運動、認知、情動などの課題に対して、どういう情報処理機構が必要であり可能なのかを理論的、トップダウン的に推論し、それを実際の脳の構造、回路、物質などの実験的、ボトムアップ的知見と対照することにより、脳のしくみを理解しようとする学問分野である」

そして、計算神経科学としての座標軸を3つ挙げている。

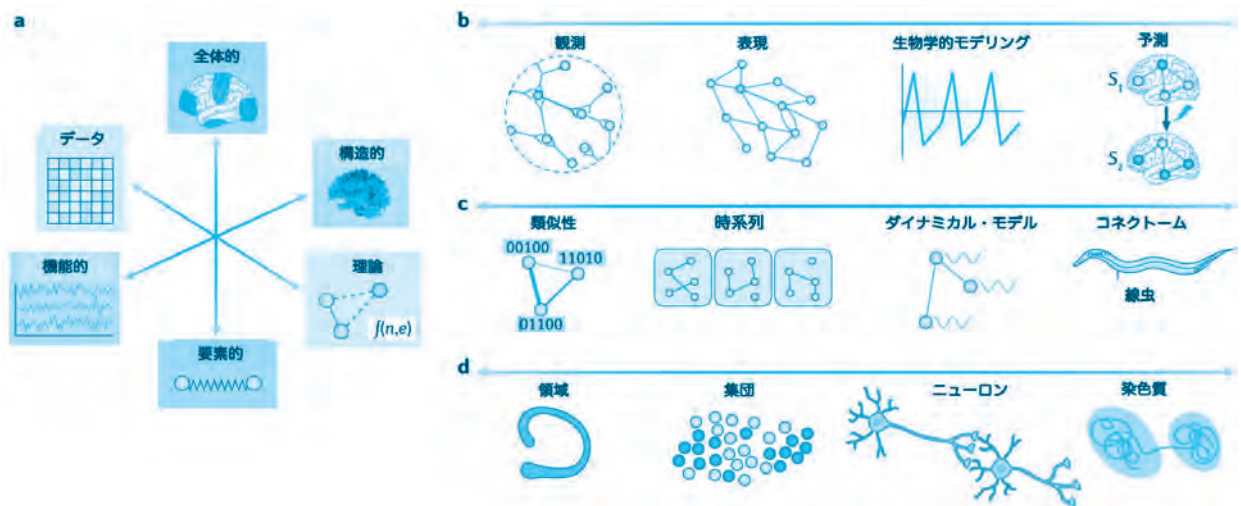
1. 空間スケール：「マクロからマイクロへ」としているが、昨今の分子生物学的アプローチの興隆をみると、ますますマイクロに向かっている傾向がうかがえる。
2. 進化と発達の時系列：「原始的なものから高度なものへ」では、進化的流れを説明しているが、「2.5.3 認知発達ロボティクスの考え方」や「2.5.4 認知発達ロボティクスにおける身体性と社会的相互作用、自己概念の発達」のような発達の側面の言及は少ない。
3. 計算の枠組み：学習のアルゴリズム：David Marr の Vision [3] で展開されている以下の3つのレベル分けは、計算神経科学の大きな指針であった。
  - (a) 情報処理の問題設定と解決指針を与える計算理論
  - (b) そのために利用可能な表現とアルゴリズム
  - (c) 実際の脳神経系のハードウェアによる実現

※1 以前は、計算論的神経科学(Computational Neuroscience)と呼ばれていたが、最近では、この言い方が定着しているようである。

現在の計算神経科学は、ビッグデータやGPGPUに代表されるリッチな計算機環境を背景に、大きく2つの流れがあると考えられる。一つは、脳活動の大規模シミュレーションで、Izhikevich and Edelman [4] による哺乳類の視床皮質のコンピューターシミュレーションが典型例である。ヒトの拡散テンソル画像 (DTI) による大域的解剖学的構造、ネコの視覚皮質の単一ニューロンの in vitro のラベリングと三次元構造再現にもとづく、複数の視床核と皮質の6層構造、そして樹状突起の多様な分岐構造を反映した22のニューロンタイプを用いて、100万ニューロン、およそ5億シナプス結合による脳神経系の動的な活動を再現し、脳波のような信号の創発を実現した。

もう一つは、ビッグデータの潮流に倣い、脳の解剖学的構造や活動に依存した各種イメージングデータをかき集め、脳の全結線構造を明らかにしようとする Connectome [5] のプロジェクトが始まっている<sup>\*2</sup>。そして、それらを対象として、脳のネットワーク構造と活動の関連を明らかにしようとする研究が計算神経科学の大きな流れの一つになっている [6]。脳に限らず、飛行機のフライトのネットワークや人間社会の構造などにも応用されているスモールワールド、リッチクラブ、スケールフリーなどの複雑ネットワークによる近似や解析が注目を浴びている<sup>\*3</sup>。このネットワーク解析に対する最新のレビューを簡単に紹介する。

■ 図1 ネットワークモデルタイプの3つの軸



出典: 文献[7]のFig.2を改編

Bassettら [7] は、“On the nature and use of models in network neuroscience” と題するレビューで最近のネットワーク神経科学のモデルに関する議論を展開している。彼女らは、多様な広がりを見せている脳のネットワーク理論は、その定義や応用範囲、そして評価などに対して若干混乱気味であるので、組織化原理を用いてこれを整理することを試みている。特に、機能と構造、データと理論、そして要素と全体の3つの軸から考察している。これらのモデル検証には、生物学のみならず哲学の分野も必要と主張している。図1にネットワークモデルタイプの3つの軸を示す。

※2 Human Connectome Project <<http://www.humanconnectomeproject.org>>

※3 <<https://ja.wikipedia.org/wiki/複雑ネットワーク>>

## 2 構成的計算神経科学のアプローチ

計算神経科学のベースとなるのは、fMRI、PET、MEG、EEGなどの各種イメージングや計測による豊富なデータと最新データ解析ツールによる脳の構造と機能(活動)に対する知見の集積である。それらは、必ずしも統一的な見解ばかりでなく、相互に矛盾するものも多々見受けられ、完全という状態から程遠い。

この大きな課題に対して、JonasとKonrad [8] は、古典的なマイクロプロセッサをモデル生物として取り上げ、神経科学の一般的なデータ解析方法が情報を処理する方法を明らかにできるかどうかを調べた。マイクロプロセッサは、論理的な流れ全体から論理ゲートを経て、トランジスタのダイナミクスまで複雑だが、すべてのレベルで構造や動作が理解できる人工情報処理システムである。すなわち真実が分かっている。そこで、マイクロプロセッサに対する様々な解析手法を適用した。そして、データ内の興味深い構造が明らかになったが、マイクロプロセッサ内の情報処理の階層を意味的に記述していなかった。これは、神経科学における現在の解析的アプローチがデータ量にかかわらず、神経システムの意味のある理解を生み出すには不十分である可能性があることを示唆している。より深い理解のために、時系列及び構造発見法の検証プラットフォームとしてマイクロプロセッサのような、既知である複雑な非線形動的システムの使用が勧められている。

結局、計算神経科学は、実はあまり頼りにならないデータや知見にもとづいている可能性があるということである。この課題に対して、完璧な解を求めるのは、非常に困難であるが、一つのアンチテーゼは、脳だけではなく脳への入出力、すなわち身体や環境の計測も必要という点である。多くの脳活動シミュレーションは脳だけに固執しており、先に紹介した IzhikevichとEdelman [4] の実験でも入力サイン波であり、出力からのフィードバックも考慮されていない。あまりに不自然である。そこで、身体からの入力や、身体への出力も考慮した計算神経科学が必要で、ここでは、これを構成的神経科学と呼ぶ。感覚器官からの入力データのリアリティ、そして運動出力から身体行動表出に至る過程の観察によるもう一つのリアリティが二重の意味で効いてくる。感覚入力器官、運動出力器官のシミュレーションは単純ではないが、少なくとも、これらの身体拘束、より正確には、「2.5 身体性とロボティクス」で示した「身体性」が脳神経系のネットワークの構造と機能の意味づけを容易にし、そのことが、認知発達ロボティクスで紹介したロボットたちの脳の設計理論につながると期待したい。

以下では、構成的神経科学の例として、胎児を扱った先駆的な研究例と、身体は単純だが、脳神経系と身体系の間関係に対する考察を扱った例を紹介し、今後の構成的神経科学の動向に注目したい。

### 2.1 胎児の発達とそのシミュレーション

近年、4次元超音波撮像などの可視化技術の進展により、胎児の様々な行動及び能力が明らかになりつつある(例えば、文献[9]の第五章など)。ただし、この時期は、自己他者未分化状態と考えられ、母胎内羊水環境で、母親の身体の内外的からの刺激としての音や光などが非明示的な他者として作用する。

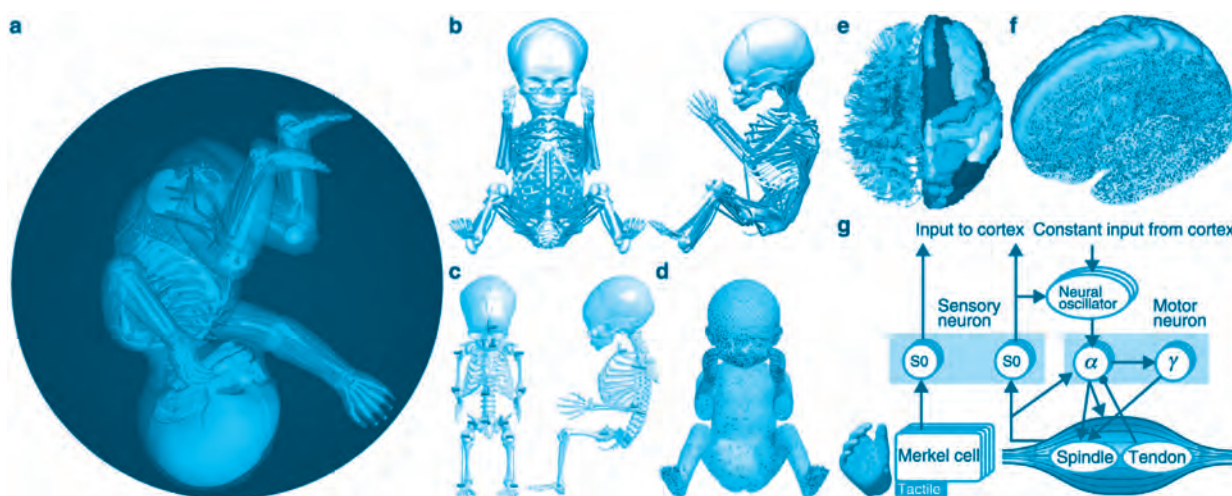
胎児の感覚の始まりとして、先にも述べたように、触覚は受精後約10週から、また視覚は18から22週の間くらいからといわれている(文献[10]のFig.1など参照[4])。身体表象が身体のクロスモデルな表現だとすると、視覚によって他者の身体を知覚する前から、自身の身体表象が触覚などの体性感覚と運動の学習からある程度獲得されると仮定しても不思議ではない。この時期は、視覚、聴覚が作動しつつも、発声や四肢の運動との明確な結び付きが薄く、それぞれが未分化、未発達な状態に

あると仮定できる。ただし、口唇周辺や手の触覚分布の高密度である点や、体内での身体の姿勢の拘束から、吸い付きなどの口唇と手の協調運動（手を口唇に近づけると口が開くなど [11]) が学習されているとみなせ、ミラーニューロンシステムの基盤として、個体の運動のライブラリが獲得され始めているとみなせる。

KuniyoshiとSangawa [13] の研究では、人の身体、神経系の生理学的知見にもとづく個々のモデルを組み合わせ、一つの赤ちゃんモデルとした。そして、このモデルを用い、母胎中の胎児の発達及び、誕生後の行動をシミュレーションし、人の運動発達の理解を目指した。学習の結果、皮質上に、筋肉ユニット配置、より一般には、体性感覚・運動マップを獲得する。この学習により母胎内では、当初ランダムであった運動が徐々に秩序化してくること、さらに誕生後、母胎外の重力場での運動は、はいはいや寝返りに似た運動が創発されたと報告されており、まさに、“Body shapes brain” [14] の典型例といえる。彼らのアプローチは、個体発達の構成的手法の基本原則と考えられる。最近では、これを起点として、脳や身体、環境のシミュレーション精度を高め、社会的行動発生原理をも含むことを狙っている [15]。そのためには、ミラーニューロンシステムのような構造が創発することが期待されるが、埋め込みとしての内的構造の基盤に加え、環境の外的構造の要件が明示されなければならない。國吉グループでの研究は、2006年の最初の論文発表 [13] では、200足らずの筋肉、すなわち、200個程度のニューロン数が、10年後の2016年に発表された論文 [12] では、260万個のニューロン、53億のシナプスコネクショを STDP (spike-timing-dependent plasticity: スパイクニューロンの活動電位タイミング依存性シナプス可塑性) 則で学習させた結果が報告されている。図2 にそのシミュレーションのモデルを示す。大規模なシミュレーションとはいえ、実際のニューロン数に比して、まだまだ少ないこと、また、子宮形状が柔らかい球状で近似されているが実際は窮屈であり、かなり異なる環境であるものの、この規模で身体との結合がなされたシミュレーションは特筆もので、今後の一つの方向性を示している。

■ 図2 ヒト胎児の筋骨格系と脳神経のモデル

(a) 子宮内の胎児身体モデル、(b) 胎児筋骨格系モデル、(c) 関節位と向き、(d) 触覚センサー配置、(e) 早産児 MRI スキャンにおける代表的な DFI と分割、(f) 皮質モデルにおけるモデルニューロン、(g) 脊髄回路モデル



出典: 文献[12]のFigure 1 より転載

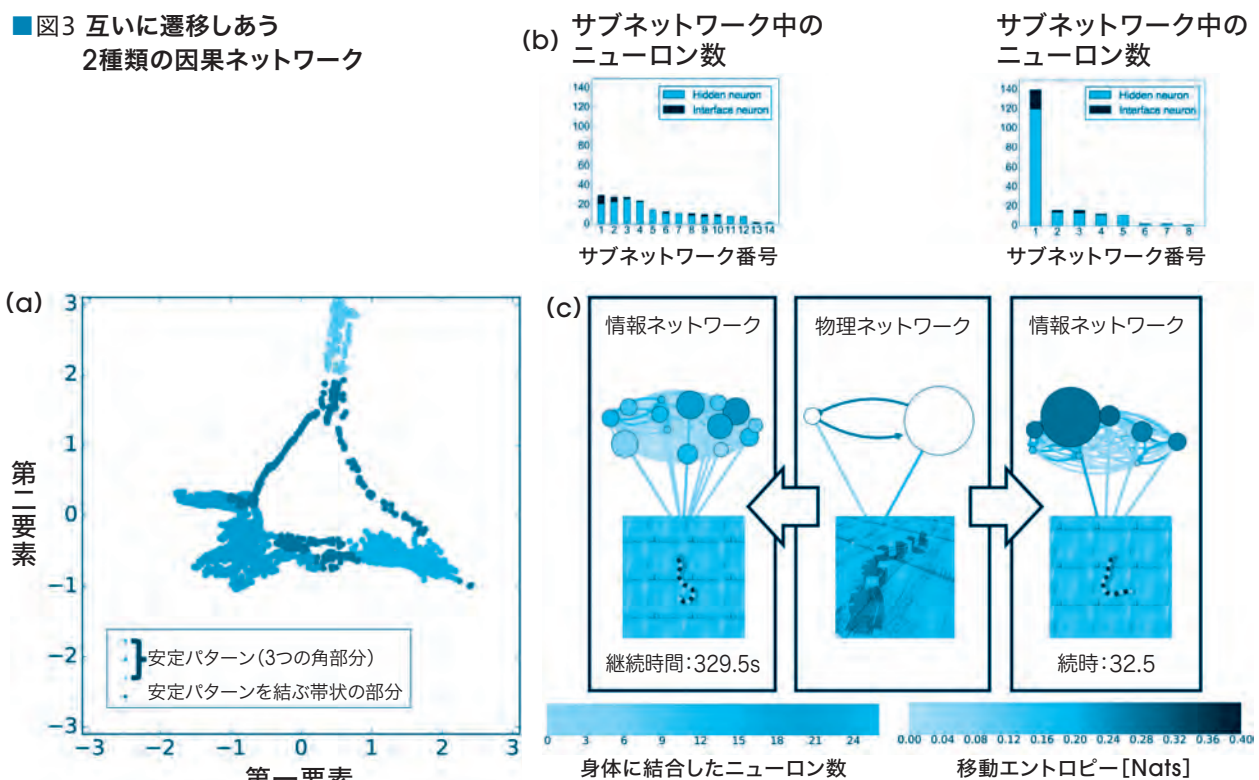
## 2.2 身体と脳神経の結合ダイナミクス

身体と環境の相互作用により、様々な行動が創発する際の感覚運動系と脳神経系がどのような関係にあるかは、構成的発達科学においての基本課題である。前節の胎児シミュレーションは魅力的だが、脳神経系と身体との関係を解析するうえでは、非常に複雑で困難である。そこで、単純な身体と脳神経系を用いたシミュレーションとその解析が行われている。

Parkら [16] は、非線形振動子のニューロンから構成される脳神経系がヘビのようなロボットの筋骨格系を通じて、環境と相互作用した際に生じるネットワーク構造について、情報の移動エントロピーをもとに解析した。初めに、各関節角の時間相関を特徴ベクトルとして、行動パターンを解析し、大まかに2つの運動パターンを抽出した。それらは、安定な行動パターン（継続時間が長い）とそれらを行き交う不安定な行動パターン（継続時間が短い）である。図3 (a) にその結果を示す。運動パターンを非線形主成分解析した場合の第一、二成分を示している。安定な行動パターン（継続時間が長い3つの角の塊）とそれらを行き交う不安定な行動パターン（安定パターンを結ぶ帯状の部分で継続時間が短い）である。

次に、それぞれの行動パターン時の神経ネットワークを調べた。最初与えられた物理的に結線されたネットワーク（解剖学的ネットワーク）が固定であるのに対し、情報の移動エントロピーの計算により、推定された運動時のネットワーク構造（以下では、機能ネットワークと呼ぶ）は、行動パターンの安定、不安定により異なるサブネットワーク構造が生じた。中央の解剖学的ネットワーク構造（図3 (c) 中央）に対し、左右の因果ネットワークが生じた（図3 (c) 左右）。左は、安定行動パターンで疎につながった（一見、密度が高そうだが移動エントロピーは低い）多数のサブネットワーク構造で、環境との結合も弱い（図3 (b) 左）。片や、右は不安定鼓動パターンで一つの大きなサブネットワークが環境と強く結びついている（図3 (b) 右）。安定行動パターンは高次元状態空間でのアトラクターに、不安定行動は、安定行動パターン間の遷移を表し、全体としてカオス遍歴の様相を呈し、環境との相互作用による神経ネットワークのダイナミクスを表している。

■ 図3 互いに遷移しあう2種類の因果ネットワーク



出典: PLOS「Chaotic itinerancy within the coupled dynamics between a physical body and neural oscillator networks」を改編

一つの憶測は、原初的な意識（不安定状態：例えば崖っぷちの歩行）・無意識（安定状態：例えば通常の歩行）に対応していないかという期待である。情報統合理論 [17] による統合情報量の計算は困難を極めるが、不安定状態のほうが安定状態よりも大きいと察せられる。

#### ◆参考文献

- [ 1 ] 浅田稔. 認知発達ロボティクスによる知の設計. 松尾豊(編), 人工知能とは, 第 6 章, pp.115–138. 近代科学社, 2016.
- [ 2 ] 銅谷賢治. 計算神経科学への招待: 脳の学習機構の理解を目指して. サイエンス社, 2007.
- [ 3 ] David Marr. Vision. W. H. Freeman and Co., 1982.
- [ 4 ] Eugene M. Izhikevich and Gerald M. Edelman. Large-scale model of mammalian thalamo- cortical systems. PNAS, Vol. 105, No. 9, pp. 3593–3598, 2008.
- [ 5 ] Olaf Sporns. Discovering the Human Connectome. The MIT Press, Cambridge, Massachusetts 02142, USA, 2012.
- [ 6 ] Olaf Sporns. Networks of the Brain. The MIT Press, Cambridge, Massachusetts 02142, USA, 2011.
- [ 7 ] Danielle S. Bassett, Perry Zurn, and Joshua I. Gold. Nature reviews Neuroscience.
- [ 8 ] Eric Jonas and Konrad Paul Kording. Could a neuroscientist understand a microprocessor? PLOS Computational Biology, Vol. 13, No. 1, pp. 1–24, 01 2017.
- [ 9 ] 明和政子. 心が芽ばえるとき. NTT 出版, 2006.
- [ 10 ] Minoru Asada, Koh Hosoda, Yasuo Kuniyoshi, Hiroshi Ishiguro, Toshio Inui, Yuichiro Yoshikawa, Masaki Ogino, and Chisato Yoshida. Cognitive developmental robotics: a survey. IEEE Transactions on Autonomous Mental Development, Vol. 1, No. 1, pp. 12–34, 2009.
- [ 11 ] 明和政子. 身体マッピング能力の起源を探る. ベビーサイエンス, Vol. 8, pp. 2–13, Dec 2008.
- [ 12 ] Yasunori Yamada, Hoshinori Kanazawa, Sho Iwasaki, Yuki Tsukahara, Osuke Iwata, Shige- hito Yamada, and Yasuo Kuniyoshi. An embodied brain model of the human foetus. Scientific Reports, Vol. 6, No. Article number: 27893, pp. 1–10, 2016.
- [ 13 ] Y. Kuniyoshi and S. Sangawa. Early motor development from partially ordered neural-body dynamics: experiments with a. cortico-spinal-musculo-skeletal model. Biol. Cybern., Vol. 95, pp. 589–605, 2006.
- [ 14 ] 国吉康夫. 赤ちゃんロボットは心を獲得できるかー構成論的科学的試みー. 日本赤ちゃん学会第 8 回学術集会, 2008.
- [ 15 ] 森裕紀, 国吉康夫. 胎児・新生児の全身筋骨格・神経系シミュレーションによる認知運動発達研究. 心理学評論, Vol. 52, No. 1, pp. 20–34, 2009.
- [ 16 ] Jihoon Park, Hiroki Mori, and Minoru Asada. Analysis of causality network from interactions between nonlinear oscillator networks and musculoskeletal system. In Late Breaking Proceedings of the European Conference on Artificial Life 2015, pp. 25–26, 2015.
- [ 17 ] Giulio Tononi and Christof Koch. Consciousness: here, there and everywhere? Phil. Trans. R. Soc. B, Vol. 370: 20140167, p. <http://dx.doi.org/10.1098/rstb.2014.0167>, 2015.

# MEMO