

# 技術動向

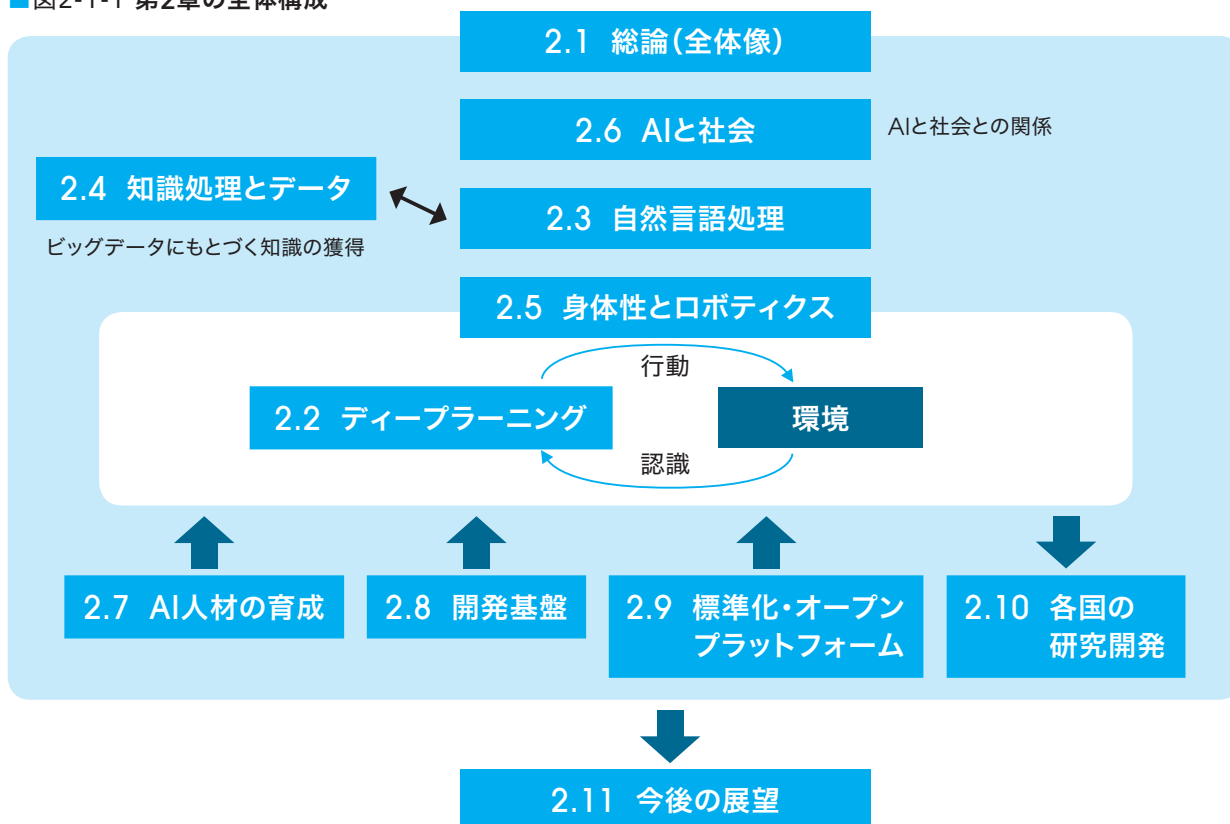
- 2.1 総論（技術の全体像、マッピング）
- 2.2 ディープラーニング
- 2.3 自然言語処理
- 2.4 知識処理とデータ
- 2.5 身体性とロボティクス
- 2.6 AIと社会
- 2.7 AI人材の育成
- 2.8 開発基盤
- 2.9 標準化・オープンプラットフォーム
- 2.10 各国の研究開発
- 2.11 今後の展望

## 技術動向

## 2.1 ▷ 総論(技術の全体像、マッピング)

本章では、AI開発に必要とされる「基盤技術」及び「技術環境」の最新動向について説明する。図2-1-1に第2章の概観を示す。各基盤技術や技術環境については、互いに関連しつつ、急速に進展している。

■図2-1-1 第2章の全体構成



基盤技術に関しては、「2.2 ディープラーニング」において第三次AIブームを引き起こしたディープラーニングの最新の動向、「2.3 自然言語処理」において自然言語処理の基本及び技術動向、「2.4 知識処理とデータ」において機械学習の学習に活用されるビッグデータやディープラーニングとの組み合わせで発展する知識処理について説明する。また、「2.5 身体性とロボティクス」においてロボットと環境とのインタラクションにより知識が創出される身体性について、「2.6 AIと社会」においてAI時代の社会の在り方などについて説明する。

技術環境に関しては、「2.7 AI人材の育成」においてAI技術者やユーザー企業の人材の育成施策や動向、「2.8 開発基盤」において学習環境やエッジにおける推論環境などの技術動向、「2.9 標準化・オープンプラットフォーム」において国際標準化やオープンソース、データの共有などの動向、「2.10 各国の研究開発」においてグランドチャレンジを含めた国内外のAI関連研究開発について説明する。最後に「2.11 今後の展望」としてまとめる。

## 2.2 ▷ ディープラーニング

本節では、ディープラーニングの定義と応用を説明し(2.2.1～2.2.4)、最新動向である、深層生成モデル、深層強化学習を説明し(2.2.5)、計算原理と実装技術についてできるだけ平易な説明を試みる(2.2.7)。

### 2.2.1 ▶ ディープラーニングの躍進

#### (1) 「眼の誕生」

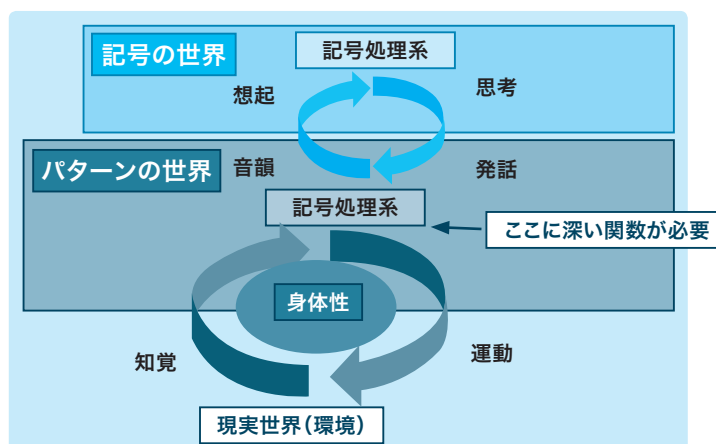
ディープラーニングは、2012年に「ILSVRC (ImageNet Large Scale Visual Recognition Challenge)」という画像認識のコンペティション<sup>\*1</sup>で大躍進を果たした。トロント大学(University of Toronto、米国)のGeoffrey Hintonらのチームがディープラーニングの技術を使って、ほかのチームがエラー率(低いほうが良い)26%台の攻防を繰り返す中、16%という脅威のエラー率を達成して圧倒的な勝利を取めた。それまでは世界中の研究者が競っても、1年で1～2%しか改善しない領域であった。その中で、10%も記録を更新し、しかも職人技の特徴抽出ではなく、自動で特徴を学習することで実現したことは驚きをもって迎えられた。

画像認識でコンピューターが人間の精度を上回ったことの意義は、いくら強調してもしすぎることはない。なぜなら、後述するように、人間の仕事の中で、「眼」を使って認識・判断している仕事はたくさんあり、それがすべて自動化・機械化できる「可能性」が出てきたからである。

#### (2) 知能の全体像

知能の全体像は、図2-2-1のようになる。まず、人間も動物も、生物はすべて環境中に生きているので、環境からのセンシングとそれに応じた行動というループが基本である。それは特定の環境に対してのみ動く、簡単な制御系でも実現できるし、より複雑な環境でロバスト<sup>\*2</sup>に動くようにも設計できる。これは、Rodney Allen Brooksが言っていた身体性<sup>\*3</sup>であり、Rolf Pfeiferの言う環境における身体性の重要性<sup>\*4</sup>である。すべての生物は環境に条件づけられた自己保存装置、あるいは再生産装置であるので、環境にその行動は埋め込まれている。

■ 図2-2-1 知能の全体像: 知覚運動系と記号系の2階建て



※1 ILSVRC2012 (ImageNet Large Scale Visual Recognition Challenge 2012)

※2 堅牢性、ここでは、外乱の影響によって変化しないという意味。

※3 「ブルックスの知能ロボット論」ロドニー・ブルックス、ISBN9784274500336

※4 「知の創成 身体性認知科学への招待」ロルフ・ファイファー、ISBN9784320120327

ところが、こうした知能の全体像において、これまでの人工知能における数十年の研究では大きな問題があった。環境中におけるパターンの処理が極めて弱かったことである。Brooksによって昆虫型ロボットが作製されても、それよりも高度なパターン処理をするものは作れなかった。特に視覚的な入力の問題は顕著であった。

ディープラーニングはこれらの問題の根っこにあった原因(特微量を自ら発見できなかったこと)を解消した。そして、そこを起点にして、様々なイノベーションが起きると予想される。順番としては、認識の問題が解決されれば、次は身体性の研究が飛躍を遂げるはずである。しばらくはロボットや産業機械の研究が最もホットな分野となる。その後、記号の研究が本格化する。今までと違って、きちんと「グラウンドした<sup>\*5</sup>」記号を使っての研究ができるからである。そして、言葉を使った意味処理が可能になり、言語からの知識獲得もできるようになる。

以降の節では、ディープラーニングの概要を説明し(2.2.2)、画像認識分野での躍進(2.2.3)や、時系列データ処理への展開(2.2.4)を紹介する。深層生成モデルや深層強化学習などの最新の話題に触れ(2.2.5)、さらに、ディープラーニングが知能の探求という意味での人工知能技術へ与えるインパクトについて述べ(2.2.6)、基本的な学習のメカニズム(2.2.7)を補足する。最後に今後の展望を紹介する(2.2.8)。

## 2.2.2 ディープラーニングとは(特長、仕組みの概要など)

ディープラーニングとは、簡単な関数を組み合わせて表現力の高い「深い関数」を作り、そのパラメーターをデータから推定する機械学習技術である。「ニューラルネットワーク」の一つと説明されることも多いが、人間の神経回路(ニューロン)を模しているかどうかはもはや重要ではない。ディープラーニングの要点は、深い関数による表現力の高さと、連鎖律(チェインルール)による勾配降下<sup>\*6</sup>にある。

本項では「ニューラルネットワーク」の構造を概観し、より広い機械学習の観点からディープラーニングの特徴を説明する。さらにディープラーニングが単に認識だけでなく、運動の習熟や、画像生成に適用され、これらが未来の予測と、プランニングにつながってゆく様子を説明する。

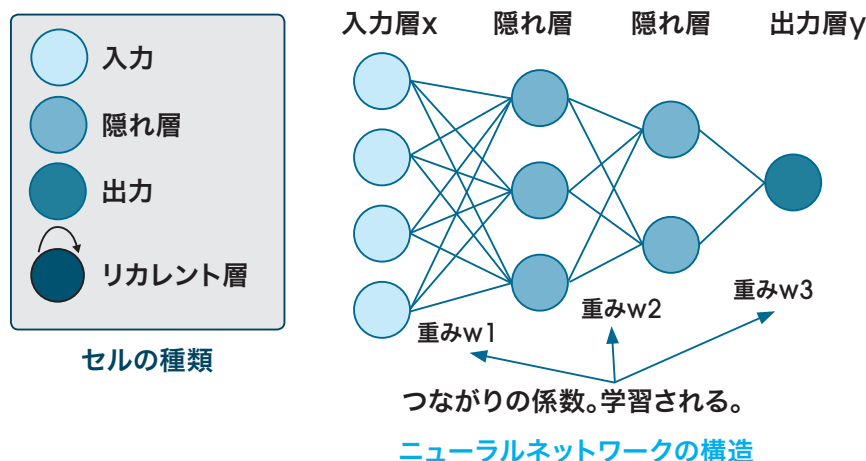
### (1) ニューラルネットワークの構造

「ニューラルネットワーク」とは、人間の脳を参考にして神経細胞(ニューロン)を多層に重ね合わせた構造を計算機の上でネットワーク上に作成したもので、神経細胞間の伝達のしやすさを調整することにより、学習を模擬する、神経細胞のような(ニューラル)ネットワークのことである。

図2-2-2は、「ニューラルネットワーク」の構造を模式的に表した図である。4層のニューラルネットワークであり、入力層 $x$ 、2層の隠れ層、出力層 $y$ で構成され、これらの間のノード(図中の円)の間に重み係数 $w_1$ 、 $w_2$ 、 $w_3$ が与えられる。学習を繰り返し、入力に対し常に所望の出力を出すように、重み係数が調整されて、所望の機能、例えば「手書きの画像を入力して、書かれている数字を判定する」、を学習する。

※5 記号システム内のシンボル(記号)がどのように実世界の意味と結び付けられるかを「シンボルグラウンディング問題」と呼ぶ。「グラウンドした」記号とは、実世界の意味が結び付けられた記号を指す。

※6 深い関数の微分は、微分の「連鎖律」を用いて計算できる。この微分値(傾き=勾配)に従って最適なパラメーター値を求めるアルゴリズムを勾配降下と呼ぶ。

■ 図2-2-2 ニューラルネットワークの構造の例<sup>※7</sup>

## (2) 機械学習とは

「機械が学習する」とは、例えば、回帰を例にとると分かりやすい。賃貸住宅に対して広さや立地条件に対して家賃のデータが大量にあるとする。広さや立地条件と家賃の間に関係性があるだろうから、何らかの近似関数を導くことができる。この近似関数が計算できると、新しい物件が出たときに、広さと立地を聞けば、だいたいの家賃を予測することができる。この予測ができる状態になることを、「コンピューターが学習した」という。

機械学習における代表的な3つの学習の枠組みは、「教師あり学習」、「教師なし学習」、そして「強化学習」である。

機械学習においては、与えられた教師データ(教師となるデータ)を、学習に用いる訓練データと、テストに用いるテストデータの2つに分けることが多い。訓練データで学習をし、その結果をテストデータで評価する。

2つに分けることにより、訓練データに対しては正しい判断ができるが、テストデータに対しては正しい判断ができないので、「過剰適合(あるいは過学習)」を避けることができる。未知のデータに対して、当てはまりをよくする能力を「汎化性能」と呼び、過学習を防ぐために、行われる処理を「正則化」と呼ぶ。

このように、正解データにもとづいて学習することを「教師あり学習」という。

一方「教師なし学習」は、正解データを用いずに、データの背後にある構造を見つけ出すものである。代表的な手法にデータの近接性からグループ分けを行うクラスタリングがある。クラスタリング以外にも、次元削減、素性学習、密度推定などが含まれる。入力そのままのものを出力するように(入力を再構成するように)構成したニューラルネットワークの一種である「オートエンコーダー」も教師なし学習の手法の一つである。

「強化学習」は、次のように定式化される。環境とインタラクションを行うエージェント<sup>※8</sup>において、ある「状態」にいるエージェントが、「行動」(環境への作用等)を選択すると、別の「状態」に移り、結果としてそれが良かった悪かったという情報(「報酬」という)が得られるとする。このとき、報酬を最大化するような行動の指針である「方策」を見つけたい。強化学習では、いろいろな行動をした後に、その情報(報酬)をもとに、どういう行動をすればよいかの「方策」を自分で学習する。

※7 表記は「The Neural Network Zoo<<http://www.asimovinstitute.org/neural-network-zoo/>>」を参考。

※8 「強化学習」の分野では、学習を行う主体のことを「エージェント」と呼ぶ。

### (3) 機械学習からディープラーニングへ

ディープラーニングは、機械学習の一つの技術である。これまでの機械学習とディープラーニングが最も異なる点は、「深い関数を使う」という点である。これまでの様々な機械学習の方法、あるいはその応用(例えば自然言語処理など)では、様々な手法が構築されたが、基本的には「浅い」ものであった(あるいはグラフィカルモデル<sup>※9</sup>などの何らかの変数間の依存関係を仮定したものであった)。

ここでは、学習の仕組みからは、「ディープラーニング系」とそれ以外の機械学習(ここでは「マシンラーニング系」と呼ぶ)に分類しよう。

マシンラーニング系のプロジェクトでは、ドメイン(対象分野)の知識を持って取り組むことが大変重要である。そして、いちばん簡単な線形モデル(重回帰)から始まって、ロジスティック回帰、サポートベクター回帰、そして様々なアンサンブル学習の手法(学習器を組み合わせる手法)、例えばランダムフォレストやXGboostなどのマシンラーニング系の手法を使ってやってみる。手法の一つとして、ディープラーニング系のアルゴリズムも適用するが、必ずしもマシンラーニングの手法より優れているというわけではない。

一方、「ディープラーニング系」は、表現力の高い深い関数を用いるため、データと計算量さえ多ければ精度が上がるという特徴を持つ(もちろんドメインの知識も重要ではある)。ただし、ディープラーニングに適したデータは限定的で、すでにデータベース化されているものよりも、画像そのものや音声そのもの、(大量にある場合の)テキストなど、生データに近いデータの扱いに優れている。ドメインの知識、一般的な背景知識等を必要とする問題に対しては、それらの知識をうまく組み込むことが難しい。

ディープラーニング系のプロジェクトでは、どうやってデータを取得するか、そのデータにどうやってアノテーション(ラベルやメタデータを与えて正解データを作ること)をするかがボトルネックになる。そして、最新の技術を使ってモデルを構築し、実際にやってみて、またデータの取り方から改良をしていくという、試行錯誤のサイクルを続けることになる。

また、ディープラーニング系と従来のマシンラーニング系との違いを最も際立たせる概念の一つとして、「エンド・トゥ・エンド(end-to-end)学習」がある。端から端を一気通貫で学習するという意味である。

ディープラーニングでは、途中をすべてニューラルネットワークとし、入力画像と判定したクラスをつなぐ。そして、正解データとの誤差から逆算して、ニューラルネットワークのパラメーターを学習する(誤差逆伝播)。従来のマシンラーニングが、データの事前処理にドメイン知識を多く必要としていたのに対し、ディープラーニングはそれほど必要としない。そういった意味で、従来の分野ごとに整理されてきた知識体系をあまり意味のないものにし、新たなパラダイムを構築しつつあるということになる。

### (4) 「ディープラーニング×強化学習」：認識から運動の習熟へ

「2.2.1(1)『眼の誕生』」で示した画像認識の次に注目すべき変化は、深層強化学習と言われる技術である。運動の習熟すなわち、機械やロボットが練習して上達するというを可能にする。

強化学習では、前述のように、学習により報酬を最大化するような行動の指針である「方策」を見

---

※9 確率変数間の依存関係を矢印で表したグラフ構造として表現したモデル。対象とする現象の裏にある構造のモデル化と、それにもとづく推論(ベイズ推論等)を確率理論に沿って行うことができる。

つけるが、その実現手段の一つとして、「こういう状況ならこういう行動を取る」というふうに、「状態」から「行動」を出力するような関数を見つける方法がある。これまでの強化学習では、関数近似と呼ばれる方法もあったが、深い関数を学習させることができなかつたため、単純な方法しか取れず、状態の数が非常に多くなると、どうしようもなかつた。さらに、人間が定義した「特徴量」を使うしなく、どうしてもあらかじめ想定した範囲内でしか「状態」を記述できなかつた。

ところが、ディープラーニングと強化学習を組み合わせる「深層強化学習」では、ディープラーニングで得られた「特徴量」を使って「状態」を定義する。これまでのやり方とは、ほんのわずかな違いだが、これだけで機械やロボットが、今までよりもずっとうまく環境の「状態」に対応できるようになった。

人間と同じように何度も試行錯誤を重ねながら、自分でゲームの攻略法を見つけていくことが模擬できる。それが可能になったのは、「状態」を認識する「画像認識」の精度がディープラーニングによって劇的に向上したからである。さらに探索による先読みと組み合わせることで、「ゲームの勝利」を「報酬」とした「方策」を学習できるようになったことがAlphaGo(碁)の躍進につながった。環境に対するロボットの一連の操作と反応をゲームとみなせば、「ゲームの勝利」とはすなわち運動を習熟することに相当する。

一例として、ディープラーニングと強化学習を組み合わせたロボット「BRETT」<sup>\*10</sup>がある。カリフォルニア大学バークレー(University of California, Berkeley)校は2015年5月に、「BRETT」が試行錯誤を重ねながら様々な動作に習熟していく動画を公開した。

この動画の衝撃を言葉に表すのは難しい。我々は、機械は「機械的な動き」しかできないし、ロボットは「ロボットのような動き」しかできないと思い込んでいる。何より、「機械的」とか「ロボットのような」という形容詞自体が、決められたとおりの動作をカクカクとした動きで行うことの意味になってしまっている。そして何回同じ動作を繰り返しても、機械やロボットは、人間や動物のように上達することはないと思い込んでいるが、この動画を見ればそれが間違いだと分かる。

もう一つ、大きな違いは、同じプログラムに別の動作を学ばせることができる点である。BRETTは目的のタスクに応じて報酬の設定をすれば、様々な動作を学習することができる。動作の種類に応じてプログラミングし直す必要はなく、同じ学習プログラムを使って、別々の動作ができるようになる。また、一旦学習した内容はコピーして、同じタイプのロボットであれば同じように動かすことができる。

## (5) 「ディープラーニング×生成モデル」：画像の生成

ディープラーニングを使うと、白黒写真に色をつけることもできる<sup>\*11</sup>。Googleの「オートドロー(AutoDraw)」は、簡単なイラストを描くと、プロが描いたようなイラストに変換してくれる<sup>\*12</sup>。これらの技術は、多くの人に衝撃を与えた。色がついた画像を白黒にすることや、イラストを簡略化することはできても、逆はできないように思える。情報量が増えているからである。ところが「オートドロー」ではそれができている。「生成モデル」では、データの生成過程をモデル化する。あるクラス  $y$  が与えられ、そして  $y$  から  $x$  が生成される確率  $P(x|y)P(y)$  を計算するので、 $x$  と  $y$  の「同時確率」を求めることでもある。ディープラーニングにより「生成モデル」を学習できるようになったの

※10 BRETT the Robot learns to put things together on his own <<https://youtu.be/JeVppkoloXs>>

※11 「ディープネットワークを用いた大域特徴と局所特徴の学習による白黒写真の自動色付け」飯塚他、2016

※12 AutoDraw<<https://www.autodraw.com/>>

で、例えば、白黒写真 (y) からカラー写真 (x) を生成するモデルを構築できる。このような能力を見ると人間特有だと思われていた「クリエイティビティ (創造性)」の一部が、人工知能でも実現されているように見える。

そもそも、人間の創造性とはいったい何であろうか？ コンピューターは本当に「創造している」と言えるのだろうか？ こうした画像を「生成」する技術はある種のエンターテインメントの用途に用いられるだけだろうか。いや、そうではない。画像生成は未来の予測につながり、次に示すように、知能の技術の進展において重要な意義をもっている。

## (6) 次のステップへ、未来の予測とプランニング

たくさんの映像からディープラーニングで学習させると、波が次にどう変化するかを予想できるようになる。そうして1枚の静止画から、その次に起こりそうな1秒間の動画を作れるようになる。こうした予測は、動物が行動計画を立てるときの基盤となる。未来を予測して、未来が自分にとって都合の良くなるように行動を選べばよいからである。

その先には、数秒程度ではなく、数分、あるいは数時間といった単位での行動計画を立てるために、記号を使ったプランニングという技術が出てくるのだが、この辺りはまだ研究者が競って研究をしているところである。

### 2.2.3 画像認識技術とその飛躍的進歩

2012年の「ILSVRC」にて、深い構造を持つニューラルネットワークが、従来手法の分類性能を大幅に上回って以来、ディープラーニングが画像認識に盛んに利用されるようになった。画像のような「空間的なデータ」に関しては、空間の隣接性を利用することができる。つまり、画像において、ある画素に遠く離れた別の画素が影響を与えることは少ない。空間は近くのものほど影響を受けるので、画像でも近くの画素のほうが影響を与える可能性が高い。この性質を使えば、近傍の画素間の関係を調べればよいので、パラメーターの数を減らすことができる。これがCNN (Convolutional Neural Network) である。本項ではディープラーニングがどのように画像を認識するのかを、CNNの具体的な例で説明する。

#### (1) CNN：画像を認識するネットワーク

CNNでは、入力に近い側から順に簡単な特徴量が学習され、それが組み合わせられてより複雑な特徴量が学習される。あらかじめイメージネットで訓練されたAlexNetやVGG-16、VGG-19、Inception、Resnetなどはよく使われるモデルである。

CNNに特徴的な処理として「畳み込み (Convolution)」と、「プーリング (Pooling)」がある。

##### ・「畳み込み (Convolution)」

画像中の小さなパッチ (例、 $N \times N$ 画素の切出し) に、特定の関数を重ね合わせて素性<sup>\*13</sup>マップを生成することに相当する。

##### ・「プーリング (Pooling)」

素性マップの局所的なパッチに対して、そのパッチに含まれる素性の「プール」(かたまり)を作り、

---

※13 対象物に対して、何を特徴とするかを「素性 (feature)」と呼ぶ。画像の場合は素性 (尖っている、滑らか等) が2次元上に配置されるので、マップになる。





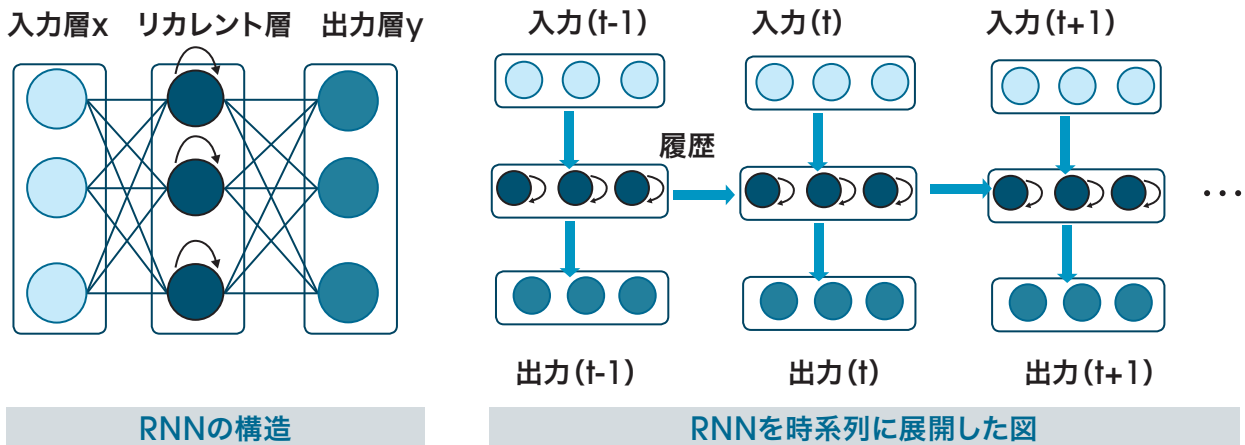
## 2.2.4 時系列データ処理への展開

音声データやテキストデータのような「時系列のデータ」に関しては、時間的な隣接性を利用できる。つまり、時系列データにおいて、ある時間的に近接した要素同士は影響を与え合う可能性が高いが、時間的に遠く離れた要素が影響を与えることは少ない。この性質を使えば、パラメーターの数を減らすことができる。これがRNN (Recurrent Neural Network) である。本項ではディープラーニングがどのように時系列を処理するのかを、RNNの具体的な例で説明し、時系列の例として自然言語処理への応用、翻訳への応用や、画像認識と組み合わせることにより、言語の意味の理解に近づくことを説明する。

### (1) RNN：時系列データを扱うネットワーク

RNNの構造は図2-2-4の左のように図示できる。右は、このRNNを時系列方向に展開される様子を図示したものである。リカレント層は1つ前の時刻の状態も入力として利用していることが分かる。

■ 図2-2-4 RNNの構造※7参照



文脈といっているものを、ニューラルネットワークの隠れ層だと考えよう。すると、出力は、この隠れ層の情報から決められることになる。そして、隠れ層の値は、1時点前の隠れ層の値と、現在の入力によって、更新されると考える。つまり、過去の文脈と新しい入力から新しい文脈が構成される。

### (2) RNN/LSTMを用いた自然言語処理

RNNの中身を分析すると、文がどのくらい長いかを学習しているユニットや、文の種類(例、疑問文か、回答文か等)、そして話題展開の階層の深さなどを学習しているユニットなど、文章の「特徴」に対応するユニットが確認できる。

こういった「特徴」を、人間がわざわざ設定しなくとも、自動的に学習することができるのが、RNNの大きな強みである。

しかし、RNNは影響が長期に及ぶときと、短期にしか及ばないときを区別できないという課題がある。LSTM(Long Short Term Memory)は時間的に長期の依存関係をモデル化しようとしたニューラルネットワークである。隠れ層のユニットに代えて、自分自身の状態を保存する「メモリセル」を導入する。不必要になったときに困るので、「忘却ゲート」を導入している。忘却ゲートがオンになると、状態を引き継がなくなる。そうでないときには、状態を引き継ぐ。これによって長期の依存関係もうまく学習でき、また、話題が変わったときには、すぐに前の状態を忘れられるようになった。

### (3) ニューラル機械翻訳 (NMT)

このようにRNNは時系列データの扱いに優れているだけでなく、柔軟に入力、出力の構造や組み合わせを取ることができることも特徴である。

例えば、時系列(sequence) データを入力・処理し、時系列データを出力するモデルは、seq2seqと呼ばれる。代表的な応用として入力と出力を異なる言語列とすると翻訳が実現できる。これは「ニューラル機械翻訳 (NMT)」と呼ばれ、従来の統計的機械翻訳 (SMT) よりも大幅に性能が向上している。

seq2seqのような変換は、画像や言語を特徴量に変換(エンコード)したうえで、画像や言語に戻して(デコードして)いるともいえる。入出力に様々な組み合わせが可能である。例えば画像をCNNで特徴量に変換しこれをRNNで文字として出力すれば、画像からその説明であるキャプションを生成する仕組みができる。

### (4) 言語の意味理解へ向けて

言語の意味理解というのは、文から画像を生成し、また画像から文を生成することであるといえる。ここで「画像」というのは、視覚的な情報を分かりやすく表現したもので、実際には、センサーとアクチュエーターの複合的な時系列情報であり、「体験」と呼ぶほうが適切かもしれない。つまり、「意味」を「理解」できるというのは、文から体験を生成し、あるいは体験から文を生成できる相互変換能力に相当する。その意味で、上記の研究は、単純なレベルの意味理解、あるいは原始的なシンボルグラウンディング<sup>\*14</sup>ができていると考えられる。

## 2.2.5 最新の技術動向

本項では、ディープラーニングの最新技術動向として、深層生成モデルと、深層強化学習を説明する。特に深層強化学習の一種であるQ学習を用いたDQN(deep Q-Network)は、AlphaGo(碁)に通じるディープラーニング応用の最前線につながる。

### (1) 深層生成モデル

ディープラーニングの発展形の一つの大きな流れが、深層生成モデルである。深層生成モデルとは、生成モデルを深くしたもの、ということであるが、まず生成モデルとは何かを説明する。

機械学習の手法の分類方法の一つとして識別モデルと生成モデルという分類がある。識別モデルとは例えば猫の画像が与えられたときに、それを「ネコ」であると「識別する」モデルである。画像認識で用いられるCNNは識別モデルである。ディープラーニングの手法ではないが、ロジスティック回帰やSVM(Support Vector Machine)などもすべて識別モデルに当たる。生成モデルというのは、「あるところにネコがいた。そのネコが画像に写り、結果として画像になった」という、データの生成過程そのものをモデル化する。生成モデルはデータ生成の過程をモデル化するので、本物に似たデータを「生成」できる。あるクラス  $y$  が与えられ(あるところにネコがいて)、そして  $y$  から  $x$  が生成される(そのネコが画像に写る)確率  $P(x|y)P(y)$  を計算する。従来の手法では、ナイーブベイズという手法が生成モデルである。

※14 記号システム内のシンボルがどのように実世界の意味と結び付けられるが明らかになった状態。

生成モデルはこれまで、グラフィカルモデルという確率変数間の依存関係を表したモデルによって実現されることが多かった。これを、深い階層のニューラルネットワークによって実現しようというのが深層生成モデルである。深層生成モデルは、創造的なデータを生成できる、複数クラスを合成したデータを生成できる、通常は検知されない異常データを検知できる、欠損値を補うことができる、などの特徴をもち、着目されている。「深層生成モデル」の代表的な手法に「変分オートエンコーダー (Variational Autoencoder ; VAE)」と「生成的敵対ネットワーク (Generative Adversarial Network ; GAN)」がある。

もともと「オートエンコーダー (Autoencoder)」は、入力そのままのものを出力するように (入力を再構成するように) 構成したものであり、隠れ層をもつニューラルネットワークで実現できる。「変分オートエンコーダー (VAE)」は、隠れ層にある潜在変数との確率的な関係をはっきりさせたものであり、例えば潜在変数を操作することで、学習したデータと似た異なるデータを生成することができる。例えばFrey Faceデータセット<sup>\*15</sup>で学習した顔に対して、潜在変数を制御することにより、怒った顔から笑った顔へ徐々に変化するデータを生成できる。

「生成的敵対ネットワーク」は、本物と同じようなレプリカを作り出そうとするジェネレーターと、生成されたレプリカが本物なのかを識別するディスクリミネーターの2つのネットワークを用意し、互いに (敵対的に) 競わせることで最終的には本物と区別がつかないレプリカを製造できるように学習を進めるというものである。そのアイデアを例えてみる。お札の偽造をする犯罪者がいるとする。そして、それを見破ろうとする銀行がいる。偽造犯は、できるだけ実際のお札に似ている偽札を作ろうとする。銀行は、どれだけ似ていても、見分けようとする。この両者がお互いに競っていくと、どんどんレベルが上がっていき、そしてついに、普通の人であれば誰も見分けられないような完璧な偽札を作ることができるようになる。

## (2) 深層強化学習

もう一つの大きな流れは、深層強化学習といわれる技術である。運動の習熟、機械やロボットが練習して上達することに関連する。

強化学習は、今まで説明した教師なし学習、教師あり学習とは異なり、「行動する主体」としての学習を念頭に置いている。長らく研究されてきた技術だが、ディープラーニングによって大きく変わりつつある。強化学習が教師あり学習などに比べて特徴的なのは、試行錯誤的な探索を行うことと、遅延報酬という2点である

「2.2.2(2) 機械学習とは」で説明したように、強化学習では、報酬を最大化するような行動の指針である「方策」を見つけることが重要であった。

こうした方策を見つけるには、大きく分けて2つの方法がある。一つは、適当に作った方策に従って行動してみて、報酬が得られるかどうかを試す。これを何度も繰り返して、試行錯誤しながら、徐々に良い方策を見つけていく方法である。これを「方策にもとづく方法」と呼ぶ。

もう一つのやり方は、状態の「価値」を計算していく方法だ。

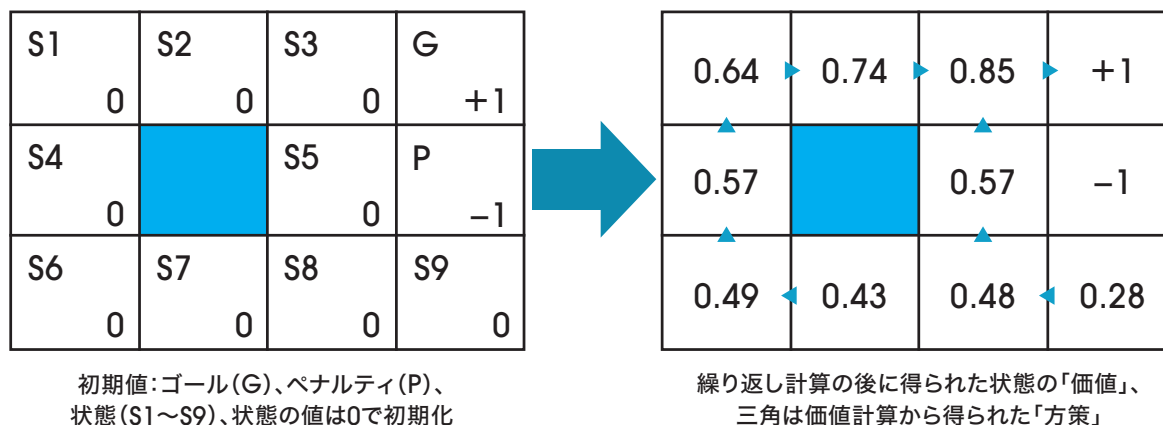
例として、3×4マスの盤上で、エージェントが、ゴールのマスG(右上)を目指して移動する例で説明する。エージェントは上下左右しか移動できないものとする(図2-2-5)。この例では通過で

---

\*15 一人の顔の表情を連続的に変化させた2,000画像ファイルのデータセット<[https://cs.nyu.edu/~roweis/data/frey\\_rawface.mat](https://cs.nyu.edu/~roweis/data/frey_rawface.mat)>

きないマスであることを表すペナルティを科すマスP(右中)もあるとする。

■ 図2-2-5 「価値反復法」により方策を計算する例



マスを状態とし、マスの「価値」の値を計算する。価値反復法(Value Iteration)というアルゴリズム<sup>※16</sup>に従えば、マスに対する「価値」を、周囲のマスの価値を使って状態の価値を再帰的に計算することができる。図の左側は状態価値の初期値を表しており、G、P以外は0が与えられている。右側は反復計算の結果を示している、右上(ゴール)の値が1なので、右上からひとつ左のところ(S3)は、1に近いくらい良い値、0.85である。逆に右下(S9)は、ペナルティに近いので悪い値0.28である。また状態価値が高くなる方向でのエージェントの動作が「方策」であり、三角にてこれを表している。以上を「価値にもとづく方法」と呼ぶ。示した例は簡単で、状態の「価値」の計算及び方策の決定はテーブル(状態×行動を軸とする)で記述できる。

しかし、実際には膨大な変数とその組み合わせに対してテーブルを作ることは容易ではない。

そこで、価値を算出するような「関数」を学習することを考える。この「関数近似」の方法も昔から研究されてきたが、これまでは、深い関数を学習させることができなかつたため、浅い関数を用いるしかなかった。それが、ディープラーニングにより、深い関数で、状態から価値(あるいは行動)の計算(すなわち方策)を学習できるようになった。

### (3) 深層強化学習の例(Q学習、DQN)

「価値にもとづく方法」の代表的な学習であるQ学習では、2つの「価値関数」を使う。ある状態の価値を表す関数を状態価値関数Vという。ある状態から将来もらえる報酬の期待値を表す際、このVを使って方策を決めるが、(将棋や囲碁などの場合を除いて)次にどの状態に遷移するか分からないことも多い。そこで、「ある状態sにおいてある行動aをすること」の報酬の期待値を表す行動価値関数Q(s,a)を用意する。

つまり、各状態での行動価値を、実際の経験からの推測値と合わせていくような操作を行う。これは、実際の経験からの推測値をY、各状態を表す変数をXと考えると、XからYを推定するような関数を求める問題に帰着する。これを、深い構造をもったニューラルネットワークで代替するのが

※16 「価値反復法」では、状態sに対して、R(s)を即時報酬関数、P(s'|s,a)を状態sから行動aで次状態s'に遷移する確率、rを割引率とすると、状態価値関数U(s)は、周りの状態の価値関数U(s')から以下の式で計算することができる。

$$U(s) \leftarrow R(s) + \gamma \max_a \sum_{s'} P(s'|s,a) U(s')$$

DQN (deep Q-Network) である。

DeepMindの「DQN」は、Q学習の行動価値関数を、ニューラルネットワークで置き換えたものである。状態  $s$  におけるアクション  $a$  の価値  $Q(s, a)$  を、画像を入力とするニューラルネットワークで計算する。アタリ社のゲームを学習させた例では、画像は、現在から4フレームまで遡り、 $84 \times 84$ の画像にクロッピングし、グレースケールにして入力する(ニューラルネットワークに必要なパラメーターを減らすため)。その後、2層のコンボリューション層、1層の全結合層があり、最終的に各アクションのQ値を出力する。エミュレーション上の49のゲームのうち29のゲームで、人間並みあるいはそれ以上のスコア(人間のエキスパートのスコアの75%以上)を出した<sup>\*17</sup>。

## 2.2.6 ディープラーニングを端緒とする人工知能の発展

AIの分野で議論されてきた様々な難問において、結局のところは、データをもとにして特徴量を抽出するところに最も大きな困難性があり、それが今、「現実的な方法で」「実際に」解けるようになっている。本項では、ディープラーニングのような「表現学習<sup>\*18</sup>」の進展により、人工知能の本質的な課題である「意味の理解」の実現が近づいていることを説明する。

例えば、階層的に履歴を集積する巧妙なRNNの構造を導入することで、高い品質の音声合成を実現できたり、教師ありのディープラーニング、強化学習、モンテカルロ木探索を巧みに組み合わせたコンピューター囲碁ソフトウェアである「AlphaGo」などが、代表的な例である。

得意なのは、画像や映像、音声などの「生に近いデータ」で、テキストの扱いも得意なものの一つに入る。しかし、これは大量にあるテキストをパターンとして学習させる統計的自然言語処理の延長線上にあり、劇的な精度の向上というより、従来と同じかそれを上回る精度を、ほとんど人間の介入なしにできるということが優れている。逆に、「生」ではないかなり加工されて抽象度の高い、「高次の特徴量」になっているデータでは、それほど既存技術に比べて効果が期待できない。

### (1) 人工知能と「身体性」

人工知能を実現するうえで、「身体性」が必要かどうかは、長らく論争的であった。

「身体性」とは、「行動体と環境との相互作用を身体が規定すること、及びその内容、環境相互作用に構造を与え、認知や行動を形成する基盤となる」<sup>\*19</sup>ことを意味する。もともと生物における知能は、生物の生存確率を上げるためであり、最終的に生物の行動に紐づかないと意味がない。その意味で、生物の知能は身体性と不可分である。

「身体性」を持たない知能の限界の例として、Terry WinogradのSHRDLU<sup>\*20</sup>を紹介する。SHRDLUは、画面の中の「積み木の世界」にブロックや円錐、球などがあり、例えば「円錐は何に支えられているか?」といったユーザーからの質問に、自然言語文で答えることができた。また「緑色の円錐を赤いブロックの上に置け」といったユーザーからの命令を実行することもできた。このように、コンピューターが少し知的に見える振る舞いをできるようになる様子を示した。しかし、SHRDLUの積み木の世界はすべて人間が設計した世界であり、ありとあらゆるお膳立てが必要であ

※17 「Human-level control through deep reinforcement learning」Nature, Feb.2015

※18 データに対して、これを特徴づける表現(特徴など)を学習すること。表現は人間に分かりやすいとは限らない。自明でない表現の例として文章における単語の特徴をベクトルとして学習するなどがある。

※19 浅田稔「身体・脳・心の理解と設計を目指す認知発達ロボティクス」,2009,計測と制御vol.48,No.1

※20 Terry Winograd, "Procedures as a Representation for Data in a Computer Program for Understanding Natural Language", MIT AI Technical Report 235, February 1971

り、そこに限界があった。積み木の世界の研究が、知能の重要な側面をとらえる素晴らしい試みであること自体は、何も間違っていない。そして、それが今、ディープラーニングを突破口に新たな展開を見せつつある。

鍵となるのが、ディープラーニングにおける「生成モデル」である。例えば、「GAN」は生成器と識別器から構成され、互いに騙そう、騙されまいとすることによって、人間があらかじめお膳立てしなくても、精度を上げることができる。

「視覚的想像 (visual imagination)」は、実際に行動しないにもかかわらず、未来の状態を予測することである。Katerina Fragkiadaki<sup>\*21</sup>は、ビリヤードの例で、「LSTM」を用いて、20フレーム先の未来まで、ボールの速度(ベクトル)を予測することができることを示した。予測ができれば、どのようなことが起こるかをシミュレーション、すなわち「想像」し、行動の計画を立てることができる。「DQN」も、アクションを挟み込んだオートエンコーダーでフレームを学習することにより、特定の行動を行うと次に何ができるかを予測する技術であるとみなせる。

このように、SHRDLUのように明示的に積み木の世界を作らなくても、ディープラーニングの生成モデルを使うことによって、その世界を描くことができるようになった。

## (2) 環境とのインタラクションで得られる「身体性」から記号に迫るアプローチ

(1)の考察から、認識能力を獲得したディープラーニングの研究が、身体性の研究に向かうのは、必然であると言える。生物(我々人間も含む)は、認識能力を活かしながら、環境とインタラクションすることで、環境の中から重要な情報(あるいは特徴量)を見つけ出しているからである。

とはいえ、現実世界は驚くほど複雑で非線形であり、現象をモデル化したシミュレーター上での「身体性」では、結局はモデル化によって重要な要素が捨象されているかもしれない。

シミュレーション上で完結する「身体性」は、どうしても「ちゃちな」ものになる。やはり、現実世界とインタラクションできる何らかの「身体」が必要と思われる。

人間も動物も、環境の中で生きているので、環境とのインタラクション、つまり、センサー(感覚器官)からインプットした情報にもとづいてアクチュエーター(運動器官)で行動をアウトプットするというループが基本である。それは特定の環境に対してのみ動くような簡単な制御系でも実現できるし、より複雑な環境でロボストに動き、また学習していくようにも設計できる。これは、Rodney Brooksの「身体性」であり、Rolf Pfeiferの「環境における身体性」である。すべての生物は、環境に条件づけられた自己保存装置、あるいは再生産装置なので、その行動は環境に埋め込まれている。

人間の場合は、この「身体性のシステム」の上に「記号のシステム」を載せている。Rodney BrooksはMarvin Minskyを「ゾウはチェスをしない」<sup>\*22</sup>と批判したが、まさにその裏返しで、人間だけが「記号」を使ってチェスというゲームができる。

## (3) 「身体性」抜きに、大量のデータから直接の概念獲得や意味理解を目指すアプローチ

「身体性」の研究があってはじめて、その先の「シンボルグラウンディング(記号接地)」の研究に進める。だから、ロボット研究における日本の優位性は大きなアドバンテージがある。しかし、

※21 Katerina Fragkiadaki, PulkitAgrawal, Sergey Levine, JitendraMalik, "Learning Visual Predictive Models of Physics for Playing Billiards"

※22 論理やアルゴリズムを突き詰めることで人工知能はチェスができるまで「知的」になったが、ゾウはチェスはしない。自然に囲まれた環境で生きる「知恵」を持っているとの趣旨。

GoogleやFacebookの研究者たちの研究のスピードを見てみると、実世界における身体性の研究を全部すっ飛ばして、大量のデータから直接、概念獲得や意味理解ができてしまう可能性もある。

Google自動翻訳がディープラーニング方式になって相当精度が上がったが、まだまだ上がるはずだ。Googleが「画像認識」や「身体性」をすっ飛ばしてこのレベルに達したということは、きちんと「身体性」を経て、本当の意味での「意味理解」ができたとしたら、自動翻訳は真に実用的なレベルに達するだろう。

#### (4) 表現学習の手段としてのディープラーニング

ここまでは、ディープラーニングに端を発する人工知能の発展の方向について述べてきたが、「深い(深い)」ネットワークを使うことは、必ずしも必然ではない。ディープラーニングは、より抽象化して「表現学習」ともいわれる。複雑な関数をデータから学習するために、今のところ「深い(深い)」ネットワークが有効である。深いことは、結局は、我々の世界を前提として学習を早くするための一つの「ズルのしかた」であって、世界にはこうした「ズルのしかた」がたくさん存在する。このズルのしかたを「prior(一般事前知識)」という。

### 2.2.7 ディープラーニングの計算原理と実装技術

本項では単純な画像認識の題材をもとに、ディープラーニングの原理と必要な実装技術について説明する。

#### (1) ディープラーニングの原理

ディープラーニングとは何かを、ひと言で説明すれば、「深い関数を使った最小二乗法」ということができる。深い関数というのは、入力Xから出力Yまで、何段階か中間の関数を介する関数である。深い関数のほうが浅い関数よりも圧倒的に表現力が高い。

例えば、100×100の画素で構成される画像にネコが写っているかを判定する問題。すなわち1万個の変数(画素)から、ネコが写っていれば1、写っていなければ0を返すような「ネコ関数」を求める問題である。

最も単純な判定式は、1万個の変数( $x_1, \dots, x_{10000}$ )に対してパラメーター( $k_0, \dots, k_{10000}$ )を用意して、以下のようなネコ関数を定義して、

$$f_{\text{ネコ}}(x) = k_0 + k_1 \times x_1 + k_2 \times x_2 + \dots + k_{10000} \times x_{10000}$$

このデータに最も当てはまりがよくなるように、 $k_0$ から $k_{10000}$ のパラメーターの値を求めればよい。これは最小二乗法で求めることができる。

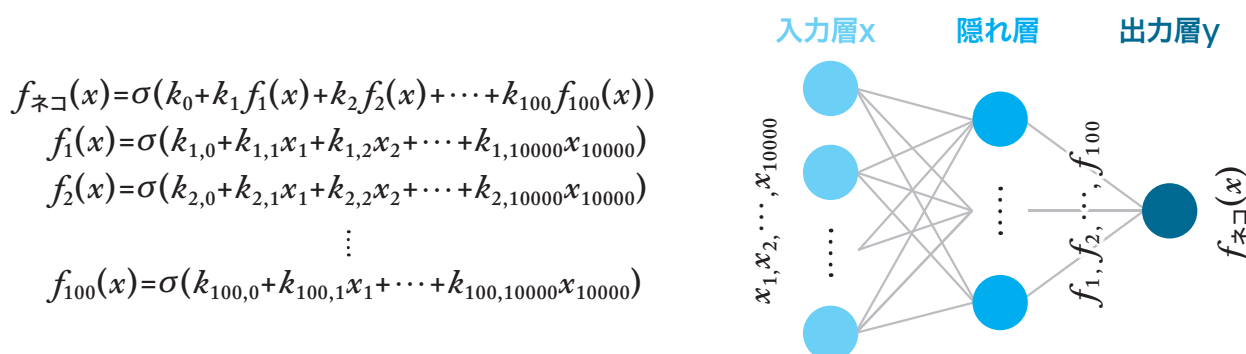
ところが「ネコ関数」はこんな単純な構造をしていない。ディープラーニングの言葉の由来となったように、「ディープ」な階層構造を導入する。ここで階層構造とは、入力データが入ってくる変数から直接ネコ関数を定義するのではなく、一旦別の関数を介して、ネコ関数を定義することである。

最初にサブとなる関数 $f_1, f_2, f_3, \dots$ を定義する。そして、これらの関数を使ってネコ関数を定義するわけである。これは2層(入出力も入れると4層)のニューラルネットワークに相当する。さらに中間的な関数を増やして階層を増やせば、3層、4層とどんどん深くしていくことができる(図2-2-6)。



ここで大切なのは、線形な関数(変数の一次式)をいくら重ねても線形にしかならないことであり、より多様なネコの形に対応するには、「ネコ関数」へ非線形な要素を入れたい。シグモイド関数 $\sigma$ は、ニューロンの働きにヒントを得ており、これをサブとなる関数にそれぞれ入れていけば、非線形の強い、様々なものを柔軟に学習できる、表現力の高い関数となる。そして、こうした関数のパラメータは最小二乗法で求めることになる。シグモイド関数に代表される非線形な要素を活性化関数とも呼ぶ。

■ 図2-2-6 階層構造に非線形性を入れた「ネコ関数」の例とニューラルネットの構造



以上をまとめると、ディープラーニングとは、「シグモイド関数等の活性化関数を使って非線形性を入れ、多層に構成した関数を使った最小二乗法」ということもできる。なお、歴史的にはシグモイド関数が使われてきたが、最近ではよりシンプルなReLU(ランプ関数)等がよく使われる。

## (2) ディープラーニングの実装技術

ディープラーニングの実装に必要な基本実装技術は以下の3つである。行列演算という並列度の高い積和演算が多用されるがゆえに、GPU(Graphics Processing Unit)との相性が良い。

### ① 行列演算による実装

多数のユニットを結ぶ重み付き有向辺(矢印の付いた辺)計算を効率的に行うために、行列演算と非線形関数の組み合わせを効率的に行う実装が望ましい。ベクトルと行列だけでなく3つ以上の軸を持つテンソルも扱えることが望ましい(例えば、センサーの信号を時間周波数解析する場合、時間 $t$ 、周波数 $f$ 、センサー番号 $i$ の3つの軸を持つデータに対して、演算を行う)。本来3次元グラフィックス処理の高速化に活用されてきたGPUがディープラーニングの高速化に寄与したのは、3次元グラフィックスにおけるポリゴンを移動させたり回転させたりするときの行列演算が、ニューラルネットワークにおける、重みの行列に別の行列を掛けて足し合わせていく「行列の積和演算」という処理と同じであったからである。

### ② 勾配の計算手法

ニューラルネットワークを用いた学習では、誤差を最小とするように重み $w$ を調整する。このとき、重み $w$ の変化でどれくらい予測値が変化するかを勾配と呼ぶ。勾配が分かれば、どれだけ重み $w$ を変化させればよいかのあたりがつく。数学的には関数の微分の計算を行う。ニューラルネットワークは数学的には重みの線形和関数や活性化関数を階層的に組み合わせた合成関数として与えられてい

るので、その微分は連鎖律(合成関数を微分するとき、その導関数がそれぞれの導関数の積で与えられるという関係式)を用いて各関数のヤコビ行列(行列版の微分に相当)の積として展開できる。

最も一般的な勾配計算法は誤差逆伝播法で、合成関数の勾配計算には自動微分と呼ばれる機能が用いられる。

### ③最適化ルーチンと効率化

勾配計算では、一つの訓練データから計算した勾配で近似して勾配計算を行う確率的勾配法が有効である。このとき、収束を早くするような様々な最適化アルゴリズムが存在することが知られている。ディープラーニングの実装ではこの確率的勾配法にもとづく最適化ルーチン、最適化のための様々な高速化技術が必要になる。

## 2.2.8 今後の展望

本項では、今まで説明したディープラーニングの発展を踏まえ、今後人工知能がどのように発展し、その中で日本が発展に寄与できる部分を明らかにするとともに、最終的に知能の原理の解明につながるであろう進展を展望する。

ディープラーニングの進展に端を発する人工知能に関して、実現可能性の高い未来をフェーズ1からフェーズ5という形でまとめた。

フェーズ1：ディープラーニングによる画像認識

フェーズ2：眼をもつ機械の誕生

フェーズ3：翻訳が実用レベルに達する

フェーズ4：言葉の意味理解・知識処理

フェーズ5：物理現象や社会現象のモデル化

フェーズ1の画像認識からフェーズ2のディープラーニングと強化学習を組み合わせた「眼をもつ機械」という流れがあり、先には翻訳、言葉の意味理解の世界が広がっている。この中で、日本が市場を取れるとしたら、ほぼフェーズ2の強化学習のところしかないということである。フェーズ3、フェーズ4の辺りは完全にG A F A (Google、Apple、Facebook、Amazon)の強い領域(つまりハードウェアが関係ない、特定の産業集積が意味をもたない)世界だからである。だからこそ、1から2のところでは日本がきちんと市場を取れるかどうか、つまり、この10年くらいの大きな変化に先導してついていけるかどうかが大変重要である。

ディープラーニングの深層生成モデルを使えば、「意味」を「理解」することができる。例えば画像キャプション生成や文からの画像生成は意味を理解しているといえる。日本語の文から体験を生成し、それを英語の文に変換するということもできるはずである。それはすなわち、意味理解を伴う機械翻訳そのものである。

ディープラーニングが、それまでの機械学習がもつ様々な問題の根っこにあった原因(特徴量を自ら発見できなかったこと)を解消し、そこを起点にして、様々なイノベーションが起きるだろうと述べた。順番としては、認識の問題が解決されれば、次は身体性の研究が飛躍を遂げるはずだ。しばらくはロボットや産業機械の研究が最もホットな分野となる。その後、記号の研究が本格化する。今までと違って、きちんと「グラウンドした」記号を使つての研究である。言葉を使った意味処理が可

能になり、言語からの知識獲得もできるようになる。

### (1) 「モデリング」の自動化

自然言語処理の分野は、多くのデータを使う統計的自然言語処理にシフトし、またデータマイニングなどの研究分野も立ち上がった。インターネット上のシステムには、こうしたアプローチが多く使われ、GoogleやFacebookなどで真っ先に活用された。おすすめ商品を提示する推薦システムも、Amazonで有名になった。こうした機械学習は極めて有用ではあったが、変数となる「素性(feature)」をどう作るか(「素性エンジニアリング」と呼ばれる)も難しい問題の一つであった。

人工知能の研究の中で難問とされるフレーム問題やシンボルグラウンディング問題、あるいは素性エンジニアリングの問題に共通するのは、現実世界の現象から何が重要かを見抜いていかにモデルを立てるかという部分である。そこが自動化できなかつたために、重要な部分の作業はどうしても人間がやらなければいけなかつたのだ。その突破口を切り拓いたのが、ディープラーニングである。ディープラーニングの研究が進めば、フェーズ1～フェーズ5それぞれで必要とされるモデル構築も加速されるはずである。

物理モデル構築の最新事例として、2018年6月にDeepMindが発表した「GQN」(Generative Query Network)がある、GQNは、2次元画像から3次元画像を生成する新技術であり、これにより平面的な写真を見て、見えていない部分を想像するかのよう、3次元空間(世界のモデル)を作り出すことができる。

### (2) 知能の原理の解明へ

人間は生存のために知能という機能を発達させた。飛ぶことに単一の原理があったように、知能にも原理があるはずである。しかも、その原理は、実は比較的単純なものに違いない。すでにディープラーニングで一部が解明されつつあり、あとは身体性や言語の意義が分かれば、知能の原理の大方は解明されるはずである。

そして、知能の原理が分かれば、これを工学的に利用することが可能である。その結果、様々な製品やサービスが生まれる。そのときに、必ずしも人間を擬す必要はない。トリと飛行機がまったく異なる形をしているように、「知能」という機能を実装した機械がヒト型ロボットである必要はないのだ。それこそが、本節で何度も強調してきた「眼をもつ機械」の実態であり、フェーズ1からフェーズ5へと続く、様々な人工知能技術の進展の姿である。

## 2.3 ▷ 自然言語処理

ある企業のサービスや製品に興味を持ち、その会社のWebページを訪れたとしよう。そこにはその企業の歴史や組織から、提供しているサービスや製品の情報まで、様々な情報が載っているだろう。しかし、知りたいことがすべて載っているとは限らず、「開店時刻は何時ですか?」など詳細について質問したくなることがある。最近では、Webページに質問を受け付ける機能が付いていることも多く、もしリアルタイムで回答が返ってくる場合は、背後でオペレーターが何人も働いているか、そうでなければchatbotが動いている。すなわち、質問に対し自動的に答えを返すアプリケーションである。このアプリケーションを支える技術が自然言語処理、すなわち人間の言語を処理する技術である。自然言語処理には機械翻訳や仮名漢字変換など多くの応用がある。とはいえ、高度な自然言語処理技術が常にchatbotに利用されているとは限らない。最も単純には、例えば「何時」という文字列が質問に含まれていれば、「営業時間は10時から19時までです」という回答を返すといったルールを用意しておくこともできる。しかし、このような単純な仕組みでは「何時頃が一番混んでいますか?」や「何時まで入店できますか?」などの質問に対し、的外れな回答をしてしまうことになる。こう考えると、「何時」だけでなく「混む」や「入店」など他の多くの単語もうまく使って回答を選びたい、質問の意味を理解させたい、事前に作成した回答リストにない回答もできるようにしたい、など様々な形でchatbotを賢くしたくなるだろう。ここでは、chatbotを賢くするという目的を通して、そのために使える自然言語処理技術を概観することにする。

### 2.3.1 文書分類

文書分類とは、与えられた文書を、あらかじめ定義された複数のカテゴリのいずれかに分類するタスクである。例えば、電子メールをスパムメールとそうでないものに分類するタスクや、ブログ記事を政治、経済、スポーツ、生活などのトピックを表すカテゴリに分類するタスクなどが分かりやすい。冒頭で述べたchatbotにおいても、例えば回答を100個用意し、それらを回答1、回答2、…、回答100というカテゴリだとみなし、さらに「何時頃が一番混んでいますか」などの質問を文書だとみなして、この文書を上記の100個のカテゴリのいずれかに分類するというアプローチが考えられる。分類されたカテゴリの回答を返せばよい。

さて、各文書は、「何時 頃 が 一 番 混 む で い る ま す か」などのように単語分割、及び基本形化されているものとする<sup>\*23</sup>。文書分類の基本的な方法においては、まず各文書をベクトルで表す。よく用いられるbag-of-words表現では、ベクトルの各要素に単語を対応させ、その単語が文書中に出現していれば1、そうでなければ0を要素の値とする。1の代わりに、その単語のその文書中での出現回数を使う場合もある。bag-of-words表現では、単語の種類数と同じ次元を持つベクトルで各文書が表現されることになる。これらの文書ベクトルを事例とし、ベクトル进行分类してくれるような機械学習手法(例えばサポートベクトルマシン)を用いる。機械学習分類器は、通常教師付き学習手法であり、すなわちラベル付き訓練データを必要とする。chatbotの例でいうと、各質問について適切な回答ラベルが付いているデータ(通常、ラベル付けは人間が行う)が、ある程度の量必要である。実際にどの程度必要なのかは、問題の難しさや分類器の複雑さによるので一概には

※23 単語分割および基本形化にはMeCabやJumanなどの形態素解析器を用いる。単語分割方法自体が研究対象となる技術であり、その分割基準、分割精度、あるいは基本形化するかどうかなどが、後段の分類精度などに少なからず影響を与えるものである。ここでは説明を割愛するが、詳細を知りたい場合は工藤の文献[3]を参照されたい。

言えない。

さて、このような単純な方法でも、難しい問題でなければかなり高精度で分類が可能である。しかし、bag-of-words表現では、各文書に各単語が出現したかどうか(あるいは何回出現したか)の情報だけが使われ、それらの単語がどういう順序で現れたかについては捨象されている。単純な方法で、わずかながら語順を考慮するものとして、bag-of-n-grams表現がある。n-gramとはn個の単語の並びで、n=2としたときは、ある2単語の組が連続して出現したかどうか(あるいはその出現回数)を考える。「何時頃が一番混んでいますか」であれば、これを「何時-頃 頃-が が-一番 一番-混む 混む-で で-いる いる-ます ます-か」とみなすことになる。これにより、例えば、「何時頃」という表現が出現しているという情報をベクトルに入れ込むことができる。さらにn=3として「何時-頃-が」などの3単語の組までいくなど、さらにnを増やしていくことも可能であるが、nが大きければいいというわけではなく、ちょうどよいnは実験的に決定することになる。

上記のn-gramのように、文書などを表現するベクトルの構成要素を素性(feature)と呼ぶ。また、機械学習手法には手をつけず、人間が知識を用いて素性を工夫して機械学習の性能を向上させることを素性エンジニアリングと呼ぶ。単純であるが実用上は効果が大きい。

文書分類の応用の一つとして、評価文書分類がある。これは、評判分析という分野におけるタスクであり、口コミなど意見が記述された文書を、好意的な意見か、あるいは批判的な意見かに分類する問題である<sup>\*24</sup>。

### 2.3.2 言語の構造解析

テキストデータから、そこで言及されている固有表現(人名、地名、組織名など)を取り出したいという状況は多く存在する。例えば、「俳優のJohn DoeさんがCMしている機種を探しています」という文には、“John Doe”という人名が含まれている。あるいは特定の機種名やメーカー名が含まれていることもあるだろう。このような固有表現を抽出する方法を考える。固有表現は一つの単語から成る場合もあるし、複数の単語から成る場合もある。上の例文が、

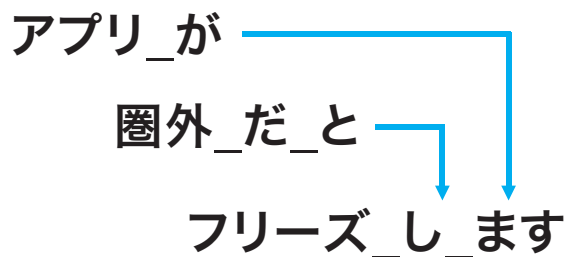
「俳優 の John Doe さんが CM している 機種 を 探 して います」  
と単語分割されているとしよう。通常、これらの各単語が、固有表現の一部であるか否かを分類することで固有表現を抽出する。もう少し詳細には、人名の最初の単語、人名の2番目以降の単語、地名の最初の単語、地名の2番目以降の単語、組織名の最初の単語、組織名の2番目以降の単語、それ以外、を分類カテゴリとする多クラス分類問題を解く。扱う固有表現の種類が増えれば、分類カテゴリ数も増える。前節の文書分類では、各文書をそれに含まれる単語で表現するなどしたが(bag-of-words表現)、固有表現抽出の場合は周辺に出現している単語(上の例のJohnなら、「俳優」、「の」、「Doe」、「さん」)などで各単語を表現する。固有表現抽出に限らず、文章の一部を抽出するタスクは、このように構成要素を分類することで実現できる。また、固有表現抽出のような抽出タスクは、言語構造解析の一種とみなされる。

もっと深い言語構造解析を紹介しよう。「アプリが、圏外だと、フリーズします」という質問をchatbotで受けたとする(図2-3-1)。前節のようにbag-of-words表現で処理してもうまく分類できるかもしれないが、質問の意味をより正確に捉えようとすると十分ではない。フリーズするのがアプリであるという情報が消えてしまうからである。このようなときに利用できるのが、構文解析で

\*24 中立的な意見も3つ目のカテゴリとして含める場合もある。

ある。すなわち、次のような構文的な依存構造を見つけることである。

■ 図2-3-1 構文的な依存構造の例



ここから、例えば「アプリが→フリーズ」のような関係の有無を文書ベクトル表現の要素として使えば、より細かい意味の把握が可能になる。このような日本語の構文的な依存構造は、係り受け構造と呼ばれ、文節（「アプリが」などは文節である）間の依存関係として定義されることが多い。文章から人々の意見、感情を分析する評判分析では、構文的な依存構造が重要になることが多い。例えば、「このアプリは非常に便利だが、デザインが悪い」という意見は、利便性という側面では良いが、デザイン性という側面では悪いという評価をしている。このように見る側面によって評価が変わることがあり、これを捉えるためには構文的依存関係を用いるのが一つの解決手段である。日本語の構文的依存構造関係を出力してくれるツールとしては、CaboCha<sup>\*25</sup>やKNP<sup>\*26</sup>などがある。この2つのツールにはそれぞれ特徴があるが、大きな違いとしては、速度面ではCaboChaに分があり、詳細な情報を与えてくれるという点ではKNPが優れている。英語の構文的依存構造解析では、文節という概念はなく、単語間の依存関係を見つける。英語の構文的依存構造解析器は多くあるが、Stanford parserなどが有名である。

実は、構文解析とは、依存構造解析だけではない。句構造解析と呼ばれるより詳細な解析もあり、英語ではこちらも発展している。句構造解析においては、“The man ordered a hot coffee”という文において、“a hot coffee”が名詞句を成し、それが“ordered”と組み合わせさせて動詞句を成し、その動詞句“ordered a hot coffee”が名詞句“The man”と組み合わせたり文を成す、といった構造を推定する。つまり、文のどこからどこまでが部分構造を成すのか、を推定する。また各部分構造は他の部分構造と組み合わせさせて上位の部分構造を成すことになる。

構文解析により、文の構造を考慮することで、文に含まれている情報をより厳密に扱うことができるようになった。この節では、もう一步意味に近づくことにする。例えば、顧客からchatbotに「電源アダプター売ってますか」という質問があったとする。この顧客は、（この店が）電源アダプターを売っているかどうかを知りたいのであるが、助詞が省略されているため字面からは「電源アダプターが」なのか「電源アダプターを」なのか「電源アダプターに」なのかが分からない。意味解析の一種である述語項構造解析では、日本語においては、各述語とそのガ格、ヲ格、ニ格となる名詞句（格要素）により内容が把握できると考え、各格の名詞句を同定することを目標とする。上記の例では助詞が省略されているが、「私は先日お店に伺い、パソコンを購入したのですが」における「購入する」のよう

\*25 CaboCha< <http://taku910.github.io/cabochoa/>>

\*26 KNP< <http://nlp.ist.i.kyoto-u.ac.jp/?KNP>>

に、ガ格の格要素である「私」が離れて出現しており格も明示されていないようなケースも非常に多い。そもそも同一の文に格要素が出現しない場合は、別の文に探しに行く必要がある。同一文書に格要素が存在しない場合さえある。つまり、このような解析は文脈解析と密接な関係がある。

ガ格、ヲ格、ニ格という表層的な格を扱うだけでなく、各要素の意味的な役割を推定する意味役割付与というタスクもある。ここで意味役割とは、例えば、その述語の動作主、対象、経験者、場所、時、目的などである。ガ格、ヲ格、ニ格という表層的な格を扱うタスクは日本語以外の言語には存在しないが、意味役割付与というタスクは日本語以外にも存在する。

表層格の述語項構造解析や意味役割付与は、述語を中心とした意味表現であるが、文全体の意味を抽象的に表現することを目的としたものとして、abstract meaning representation (AMR) という記述の枠組みも存在する。

また、意味解析と関連する重要な概念としてグラウンディングというものがある。これは、言語表現とそれが指す実体とを結び付けることである。特に、文章中に出現した固有名を、それが指す人物や場所などに結び付けることを目指すことが多い。通常、人物リストや場所リストを事前に用意しておき、文章中の各固有名がそれらのリスト中のどのエントリに対応するかを決定する、という形で問題を解く。例えば、「渋谷でオバマさんを見ました」という文において、「渋谷」が東京都にある渋谷であり、「オバマさん」が「バラク・オバマ元大統領」であることを認識する<sup>\*27</sup>。

### 2.3.3 言語資源

これまで、主に機械学習にもとづく手法などを紹介してきたが、シソーラス(類義語辞典)のように、ある単語とある単語が意味的に類似していることを示す言語資源が存在する。英語の類義語情報を含む言語資源としてはWordNet<sup>\*28</sup>が圧倒的に有名である。WordNetは、類義語関係、上位語一下位語関係、構成要素一被構成要素関係など、単語間の関係を提供している。互いに類義語関係にある単語からなる集合をsynsetと呼ぶ。すなわち、同一のsynsetに属する単語はすべておおよそ同じ意味を持つものとされている<sup>\*29</sup>。これを用いることで、文書表現としてbag-of-words表現の代わりにbag-of-synsets表現(各synsetが文書内で出現したか否か)を使うことなども考えられる。

日本語のシソーラスも紹介しよう。WordNetを日本語へ移植した日本語WordNetも存在するが、もともと日本語に対して構築された分類語彙表<sup>\*30</sup>、日本語語彙体系<sup>\*31</sup>、EDR電子化辞書<sup>\*32</sup>などがよく知られている。

その他、様々な言語資源やコーパスが、言語資源協会<sup>\*33</sup>、ALAGIN<sup>\*34</sup>、情報学研究データリポジトリ<sup>\*35</sup>などで公開されている。

近年目覚ましい発展を見せたのが、単語分散表現である。これは、コーパスから機械学習により得られた単語のベクトル表現である。すなわち、各単語が、例えば100次元、500次元、1000次元などの実数値ベクトルで表現される。コーパス内での振る舞いが似ている単語同士は互いに似たベク

※27 もちろん、この例文だけからは何が正解であるかは分からない。

※28 英語の概念辞書 <<https://wordnet.princeton.edu/>>

※29 細かい意味の違いは存在する。

※30 国立国語研究所コーパス開発センター <[http://pj.ninjal.ac.jp/corpus\\_center/goihyo.html](http://pj.ninjal.ac.jp/corpus_center/goihyo.html)>

※31 日本語語彙大系 <<http://www.kecl.ntt.co.jp/icl/lirg/resources/GoiTaikei/>>

※32 EDR電子化辞書 <<http://www2.nict.go.jp/ipp/EDR/JPN/Intro.html>>

※33 言語資源カタログ <<http://www.gsk.or.jp/catalog/>>

※34 ALAGIN <<https://alaginrc.nict.go.jp/>>

※35 NII情報学研究データリポジトリ <<https://www.nii.ac.jp/dsc/idr/>>

トルで表現される。これにより、例えば「開店時刻」、「開店時間」、「営業時間」などについて、それぞれの単語分散表現間の余弦値(コサイン)を計算し、その値が高ければ単語同士が類似していると判断できることになる<sup>※36</sup>。こうすることで、「開店時刻」を問う質問に対して用意した回答を、「開店時間」を問われたときにも利用することができる。単語をベクトルで表現すること自体は以前から行われていたが、大量のデータを使って、表現そのものを最適化することにより、容易に高精度のベクトル表現が得られるようになった。単語分散表現を獲得するためのツールとして最も知られているのが、word2vec<sup>※37</sup>である。単語分散表現は、それ自体興味深い性質を持っているが、応用上に非常に重要である。前述のように、単語分散表現を求める多くの手法は、コーパスさえあれば機能する。つまり、例えばインターネット上の大量のテキストコーパスを用いて単語分散表現を求めると、非常に多くの単語の分散表現が得られる。これを他のタスクにおいて単語を表現することに使うことができる。例えば前節の固有表現抽出では、分類したい単語の周辺の単語を用いてベクトル表現を作る。この際、周辺にある単語が出現したか否かでなく、その単語の分散表現を用いることができる。こうすることで、訓練データに出現しなかった単語でも、その分散表現さえあれば、手がかりとして利用できることになる。

コーパスから自然言語処理技術を用いて獲得された単語分散表現は、独立した言語資源として様々な場面に利用されている。このように、単語分散表現に限らず、多くの言語資源が自然言語処理技術を用いて構築され、再び自然言語処理に利用されている。

### 2.3.4 テキスト生成を伴う研究課題

これまでは、chatbotに対して問われた質問に様々な言語解析を施すことで、あらかじめ用意した回答集合のいずれかから適切なものを選択する、という枠組みを念頭に置いていた。しかし、これでは限られた種類の回答しか返せず、システムとしての柔軟性に欠ける。これに対して、古くからテキスト生成技術の開発がなされてきた。回答そのものでなく、回答のテンプレートを用意し、必要な情報をテンプレートに埋め込んで回答を生成する方法も以前から存在する。例えば、曜日によって開店時刻が異なる場合など、「本日の開店時刻は、□時□分です。ご来店をお待ちしております。」のようなテンプレートを用意し、その日の開店時刻で□部分を埋めて回答とする。開店時刻が異なる多くの営業店を持つ場合などは有効だろう。

テキスト生成を伴うタスクは応答生成に限らない。機械翻訳、文書要約など、多くの応用タスクはテキスト生成を伴う。機械翻訳は自然言語処理において、非常に重要な応用タスクであり、多くの研究者がこれに関わっている。その機械翻訳で発展したモデルが、統計的機械翻訳モデルである。これは、統計を用いた機械翻訳モデルの総称のように聞こえるが、そうではなく、特定のモデルを指す名称である。統計的機械翻訳モデルは、言語モデルと翻訳モデルから成る。言語モデルは出力文がその言語の文として自然であるかを捉え、翻訳モデルは出力文が入力文の意味に対応しているかを捉える。直観的には、例えば日英翻訳の翻訳モデルでは、各日本語の単語が、ある英語の単語に翻訳される確率を学習データから計算しておき、これを用いて全体として出力の英語文が入力の日本語文に対応しているかを捉える。その際に、単語の並べ替えなども考慮できるモデルになっている。詳細は文献[4][5]を参照していただきたい。

※36 そもそも、「開店時刻」、「開店時間」、「営業時間」が、単語分割においてそれぞれ1単語として認識される必要があるが、これは単語分割の基準や精度に依存する問題である。

※37 word2vec<<https://code.google.com/archive/p/word2vec>>



近年のニューラルネットワーク（あるいはディープラーニング）の発展は、自然言語処理分野にも大きな影響を与えている。前項で紹介した単語分散表現のような表現方法の発展がその一つであり、また自然言語処理の各要素技術についても精度が向上したという報告が多くなされている。しかしその中で特に大きな影響を受けているのがテキスト生成技術である。ニューラルネットワークにもとづく機械翻訳のモデルとしてsequence-to-sequenceモデルが提案され、またたく間に改良が施され、非常に高い精度が報告されるまでになっている。このモデルは、機械翻訳にとどまらず、文要約、応答生成などにも適用されている。以下では、機械翻訳を念頭に置き、このモデルを説明する。

■図2-3-2 sequence-to-sequenceモデル

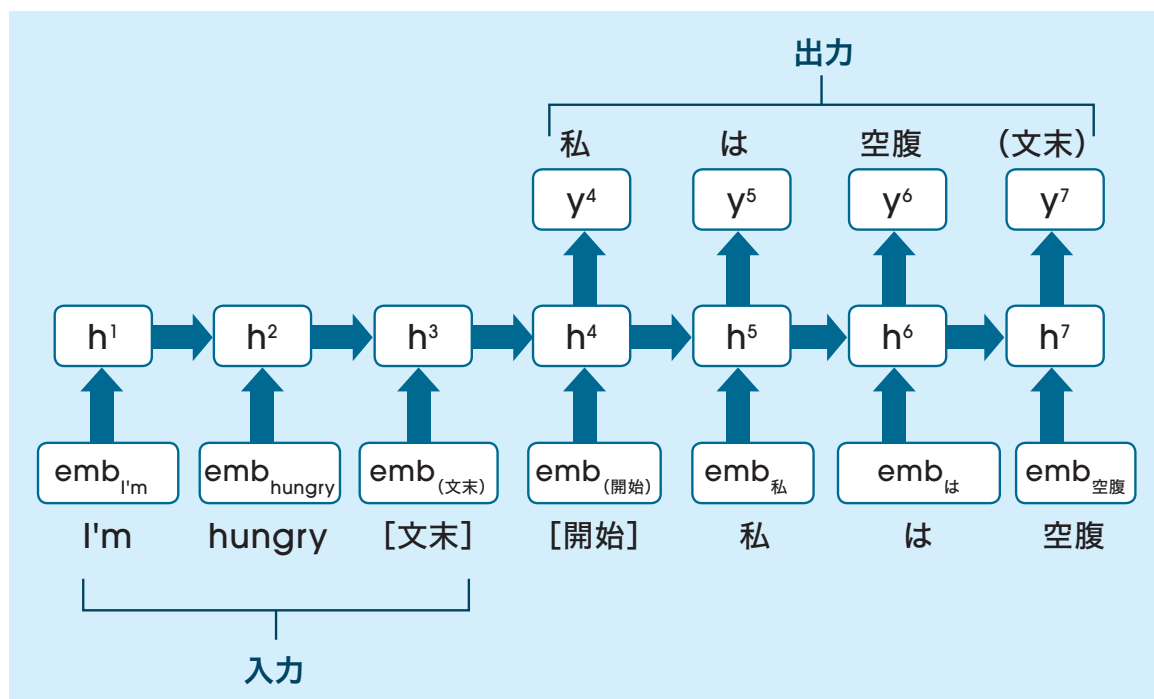


図2-3-2に示すように、「I'm hungry」を「私は空腹」に訳す場合を例にとって、sequence-to-sequenceモデルを説明する。まず、入力の最初の単語である「I'm」がベクトル表現 $emb_{I'm}$ に変換され、これがさらに行列により変換されたうえで $\tanh$ などの非線形関数を経て、新たなベクトル表現 $h^1$ に変換される。この $h^1$ が隠れ層の最初の状態である。次に、同様に「hungry」がベクトル表現 $emb_{hungry}$ に変換され、さらに行列により変換される。この変換結果は、隠れ層 $h^1$ が行列により変換された結果と足し合わされたうえで、非線形関数を経て、新たなベクトル表現 $h^2$ に変換される。これが入力文の文末まで繰り返され、得られた隠れ層の最終状態（この場合は $h^3$ ）に、入力文の意味が表現されていると考える。出力開始記号「開始」を入力し、先ほどと同様の計算を行う。ただし今度は得られた隠れ層 $h^4$ から単語の生成確率分布を計算する。確率最大の単語（この場合は「私」）を出力し、この単語をさらに次の入力とする。このように単語を一つずつ生成していき、文末記号が生成されるまで繰り返す。これが基本的なsequence-to-sequenceモデルの挙動である。

学習の際は、各単語の分散表現や、変換のための行列などを求めることになる。word2vecなどで別のコーパスから求めた分散表現を使うこともできる。sequence-to-sequenceモデルに対しては、注視機構を追加するなど、様々な精度向上の試みがなされている[6]。chatbotに利用可能な質問に対する応答生成の場合は、入力が質問、出力がその応答となる。ただし、訓練用の質問応答対データは異なる営業所や店舗のものである可能性もあり、その場合間違った情報を答えてしまう可能

性がある。そのような場合は、営業時間、店舗の場所、商品の値段など、外部データを参照する機構を組み込む必要があるだろう。

機械翻訳、文要約、質問に対する応答生成などの場合は、出力だけでなく入力も文であった。しかし、入力側に適切なニューラルネットワークを用いることで、画像を入力として文を生成する画像キャプション生成や、数値データを入力として説明テキストを生成するdata-to-text生成などにも同様のモデルが適用可能である。このようにsequence-to-sequenceモデルを一般化した場合、encoder-decoderモデルと呼ばれる。

#### ◆参考文献

- [ 1 ] T. Mikolov et al., "Distributed representations of words and phrases and their compositionality" Neural Information Processing Systems, pp.3111-3119.
- [ 2 ] 岡崎直観「<特集>ニューラルネットワーク研究のフロンティア「言語処理における分散表現学習のフロンティア」『人工知能(人工知能学会誌)』vol.31 No.2, 2016.3, pp.189-201.
- [ 3 ] 工藤拓、「形態素解析の理論と実装」、2018、近代科学
- [ 4 ] 渡辺太郎、今村賢治、賀沢秀人、Graham Neubig、中澤敏明「機械翻訳」、2014年1月、コロナ社
- [ 5 ] Philipp Koehn, "Statistical Machine Translation", 2009年12月, Cambridge University Press.
- [ 6 ] Thang Luong, Hieu Pham, and Christopher D. Manning.  
Effective approaches to attention-based neural machine translation. In EMNLP, pp. 1412-1421, 2015

## 2.4 ▷ 知識処理とデータ

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

本節ではAIの背景となるデータと知識に関して、以下について記す。

- ・ AIとデータ及び知識の関わりの経時的俯瞰
- ・ ビッグデータの状況と課題
- ・ 知識を巡る状況

### 2.4.1 AIとデータ及び知識の関わりの経時的俯瞰

コンピューターによって人間の知的機能の実現を目指す人工知能(AI)の考えの起源をたどると、万能計算機械としてのコンピューターの原理とほぼ同じ1930年代に芽生えたといえる。コンピューターの基礎を構築した偉大な科学者(Aran Turing、John von Neumann)などが、AIという言葉が生まれる前のコンピューターによる知能についても思考を巡らせている。

AI(Artificial Intelligence)という名称が生まれ、AIという研究分野が陽に形成されたのは、1956年のダートマス大学(Dartmouth College、米国ニューハンプシャー州)で開かれた研究集会からである。

その後、1950年代後半から1960年代にかけてのAI研究(第1次AIブーム期)は、主としてゲームなど規定しやすく結果の評価も容易な、いわゆる閉じた世界の問題を対象にして、知能発現の源は探索にあるとし、探索の効率化が中心課題とされてきた。探索の組み合わせ的爆発を防ぎ、効率化を図るために、ヒューリスティックス(強いて日本語訳すると発見的知識)の利用が探求された。

このヒューリスティックスは知識の一種であるが、2種の形態がある。AIの代表的な最適経路(最小コスト経路)探索法であるA\*(エースター)探索で用いられるヒューリスティック関数はゴールまでのコストの(内輪の)推定値であり、系統的な知識と言える。一方、ある状態において探索で次に進む筋道を直接的にガイドする知識(経験的知識とも称され、多くの場合ルール型)も重要な役割を担い、知識重視の次のAIの第2期で中心的な役割を果たすことになる。

1970年代後半に“Knowledge is power”といったように知識の重要性が唱えられ、実用的AI指向の「知識工学」が提唱され、1980年代にかけて知識ベースにもとづく多くのエキスパートシステムが作成された(第2次AIブーム期)。この時の知識は対象とする問題解決に向けての推論に役立つ(専門的あるいは経験的な)知識の集合体であり、形式としては大部分ルール型であった。一部に階層構造を持つフレーム型や、論理型、制約型も用いられた。知識の形式はともかくとして、この1980年代の知識の大半は(専門家へのインタビューや資料参照を介して)人が書き下したものであり、今日のビッグデータ時代に比べて小規模であり、スケールアップすることが難しいのが大きな問題であった。

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

ところで、知能(Intelligence)には、①記号化された知識とその(推論/思考による)活用の側面と②知覚系に相当するパターンの認知(あるいは認識/識別)による側面とがある(他にもいくつかの側面があるが)。以上に述べたAIは主に①の(記号化された)知識の活用の側面であるのに対し、②の知能の側面はAI分野とも言えるが、やや異なる面があるパターン認識の分野として研究、発展してきた経緯がある。今日の深層ニューラルネットワーク(Deep Neural Networks; DNN)の関連で、その経緯を振り返っておく。

人工ニューラルネットワーク(ANN: Artificial NN)のモデルは最初に1943年にWarren S. McCullochとWalter J. Pittsにより提案された。その後、1958年にFrank Rosenblattにより3層(入力層、中間層、出力層)から成るパーセプトロン(Perceptron)が提案され、入力パターン識別の学習ができることが示された。第2次AIブーム期の中の1986年に発刊されたJ. L. McClelland and D. E. Rumelhart(著)“Parallel Distributed Processing”(MIT Press; 略してPDP本; 邦訳[PDPモデル] <甘利俊一、他訳、産業図書、1988>)は、記号処理とは異なるニューラルネットワーク(NN)による分散表現であるコネクショニズムを喚起し、小ブームになった。バックプロパゲーション(誤差逆伝播法)は、それ以前から研究はあったが、この時以降にNNの学習法として定着した(後のDNNにおいても変わらずに主要な学習法である)。この当時はDNNの学習は可能になっておらず大ブームになることはなかったが、Geoffrey E. Hintonはこのころから活躍しており(上記PDP本にも共著論文あり)、後にDNNの学習を可能にする中心的役割を果たすことにつながっている。

訓練データにより学習されたDNNも(代表的にパターンを識別する)一種の知識と言えるが、上記①の記号化された知識とは異なる形態となり、使われ方も異なる。記号化された知識は多くの場合、要素的知識の集合体の形態(知識ベースに蓄えられる)となるが、学習済みDNNは不可分で一体化された形態で、一体として代表的には入力パターンの識別/認識するのに使われる。人間の脳機能に対応づけると、DNNは感覚器の部分で入力パターンを認知するのに用いられ、記号化あるいは言語化された知識は概念レベルでの思考/推論に用いられると言えよう。

データと知識のかかわりの歴史的俯瞰の話題に戻ると、情報化の進展につれて社会のデータ量は増加の一途をたどっているが、1990年代中ごろからのWebの登場と拡大、2000年代にはIoTやCPS(Cyber Physical System)などと呼ばれる物理空間とサイバー空間の情報を連携、融合させる動きが顕著になってきた。これにより物理空間も含めて多種大量のデータがもたらされ、新サービスに活用されるようになってきている。そして2000年代中ごろより、ビッグデータの時代と言われるようになってきた。ビッグデータを分析し価値を見出す方法として、統計的手法を主とするデータサイエンスによる活用が図られたが(担当者はデータサイエンティストと呼ばれる)、2012年のDNNの興隆以後はDNNも主力のアプローチになった。DNNの学習には大量のデータを要することから、ビッグデータとAIの関りは深くなった。

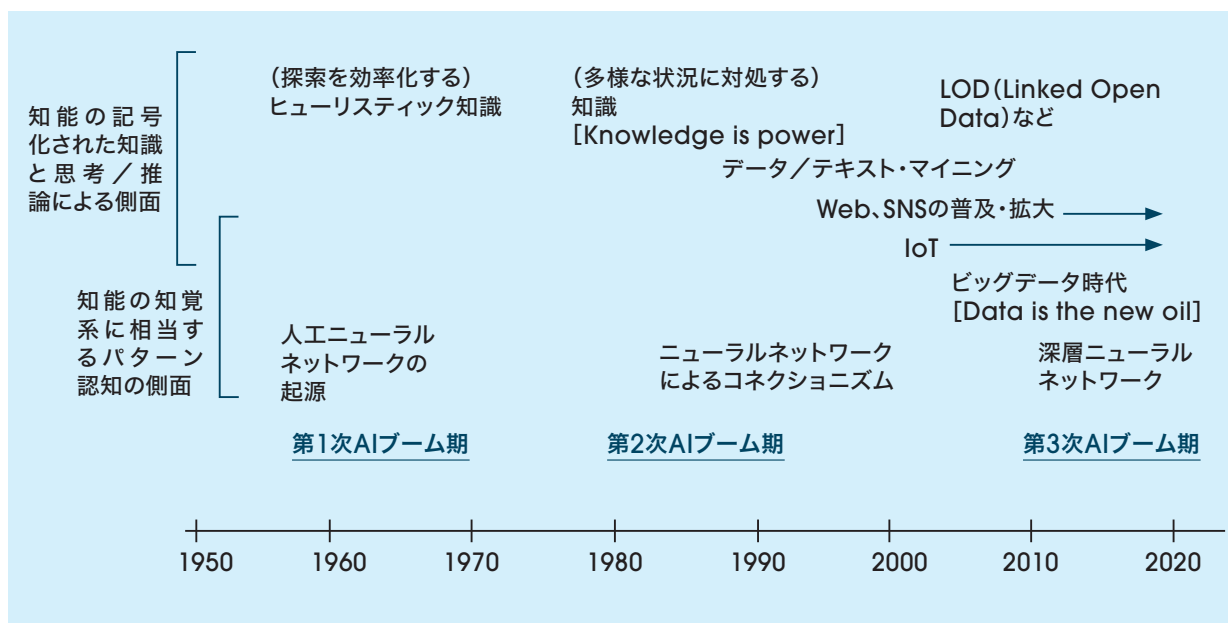
データサイエンスとAIは技術的に異なる領域と見なされるが、パターンの識別/認識の技術は両者に共にかかわって扱われることが多い(ただし、DNNはデータサイエンスに含まれることは通常なく、AIの範疇として扱われる)。

DNNを中心として第3次AIブーム期となったのであるが、この第3期のAIではビッグデータの利活用が大きな関心事になり、今日に至っている。今日、データの重要性は“Data is the fuel for AI”、“Data is the new oil”などとよく喧伝される。また、データ駆動型意思決定、データ駆動型経営、データ駆動型社会といったように、データの役割は今後増大すると予想されている。

DNNの新技术と、ビッグデータによるDNN学習(前述のように知識の生成でもある)とその活用が注目されることが多い現在であり、記号的知識の面は相対的に注目度が下がっている感がないでもないが、知能の側面として必要であり、重要性は変わらない。身近な領域として、チャットボットやスマートスピーカー(Amazon Echo、Google Homeなど)の対話では、記号的知識の利用は不可欠である。次の世代のAIは(人間の知能のように)、①記号化された知識と、②DNNを中心とするパターン認知機能を合わせて利用する枠組みが一つの重要な方向となろう。

図2-4-1は以上に記したAIをデータと知識の観点で経時的に俯瞰した概観を示している。以下にはビッグデータと知識に分けて、それぞれの状況を記すことにする。

■ 図2-4-1 AIをデータと知識の観点で経時的に俯瞰し、各時期の中心的技術項目の概観



## 2.4.2 ビッグデータの状況と課題

データ量は増加の一途をたどっており、IDCの調査によると全世界でデジタルデータの年間生成量は、2016年には16ZB(ゼットバイト)<sup>※38</sup>あったが、2025年には10倍の163ZBになると見積もられている(監視カメラ・データのように、エッジで処理され圧縮された結果だけデータセンターに送られる形態も増えることから、このようなデータも含めればさらに増えることになる)。

2000年代中ごろからビッグデータ時代と称されるようになったが、このビッグデータは単に量が大きいというだけでなく、次のような3つのVで表される特徴をもつ。

- 1) Volume (データ量) : 量が大
- 2) Velocity (速度/更新頻度) : データの生成・収集・分析の速度、頻度が高い
- 3) Variety (多様性) : 構造化データだけでなく、テキスト/画像・映像/音声/センサー情報等の多様な非構造化データも含む

※38 1GB(ギガバイト)=10<sup>9</sup>B, 1TB(テラバイト)=10<sup>12</sup>B, 1PB(ペタバイト)=10<sup>15</sup>B, 1EB(エクサバイト)=10<sup>18</sup>B, 1ZB(ゼットバイト)=10<sup>21</sup>B, 1YB(ヨタバイト)=10<sup>24</sup>B.

社会、個人生活、産業のあらゆる活動からデータは生み出されるのであるが、現在、利活用が図られているのは(必ずしも網羅しているわけではないが)以下のような領域である。

健康・医療、移動・交通、金融・保険、小売、ターゲット広告、流通・運輸、サービス業、製造業、一次産業、バイオ・製薬、化学・材料、旅行、教育サービス、エネルギー・インフラ、防災・防犯、公共・行政

これらの領域とはいくぶん趣を異にするが、データ中心科学で扱われるビッグデータ(気象や天文、バイオなど)もある。

Web情報、一部SNS、後述するオープンデータのように、オープンにアクセス可能なパブリック・データも増大しているが、世界のデータ総量の70～80%は企業が保有するプライベート・データであるといわれている。

企業の保有するデータは、

- a-1) インターナル・データ(社内のオペレーション・データ)
  - a-2) エクスターナル・データ(社外の顧客データ)
- に大別できる。

別の観点で、次のような分類の意識も必要である。

- b-1) エssenシャル・データ(自社の競争力として外に出せないデータ)
- b-2) 業界シェア・データ(業界等でシェアすることにより価値を高められるデータ)
- b-3) パブリック・データ(公共の利益に役立てるべきデータ)

また、Velocity(速度/更新頻度)の点から、次の分類もある。

- c-1) 変化のない/少ないデータ
- c-2) 更新頻度が高いデータ

このようなデータからDNNをはじめとする機械学習によりパターンや状況の認知/判定を可能にし、価値を生み出すことになる。それを可能にするためには、まずソースとなるデータが必要である。好ましいのは日々のオペレーションから大量データを収集でき、利用できることである。Google(Alphabet)、Facebook、Amazon(いわゆる国際的プラットフォーマー)はこのようにして大量のデータを収集し、ターゲット広告やリコメンデーションのようなビジネスに活用している(最近ではEUを中心に「個人データは本人の同意の下で収集し、利用する」という意識が高まってきており、このような動きは一定のブレーキがかかる動きも生まれている)。中国のAlibaba、Tencent、Baiduも同様にして大量データを収集、活用している。このような米国、中国のプラットフォーマーの規模と比べると、日本では匹敵する規模のインターネット/Webサービスの主体はなく、残念ながら量的なデータ収集力で対抗できない状態である。

一方、上に挙げたようないろいろな社会や産業活動において生まれたデータは記録されている(記録はされていても必ずしも機械学習に適するような形式や内容でないことも多く、利活用には苦勞す

る場合が多い)。今後は「データは価値を生む資産」との考えの下、収集と利活用を意識した蓄積を図るようになる必要がある。

自社あるいは自組織でのサービス／活動からデータ収集できない場合には、他所から収集する必要があり、およそ次のような方法が取られる(表2-4-1)。

■表2-4-1 自組織以外からのデータ収集の方法

利用の分類	備考
オープンデータの利用	基礎的参照データとしては有用であっても、他と差別化した特徴を出すには、これだけでは不十分なことが多い
Web等から独自に収集	Webのクローリング、スクレイピング等による
SNSデータ等の購入	Twitterデータなど、有料のものが多い
他企業からの提供	双方あるいは複数社間での協業によるデータ・シェアの枠組みも追求されている

研究用途に企業等から提供されているデータセットがあり、例えば国立情報学研究所(NII)では以下のようなデータセットを研究用に公開している<sup>\*39</sup>。

- Yahoo!データセット(yahoo!知恵袋データ)
- 楽天データセット(楽天市場商品データとレビューデータ、楽天トラベルの施設データとレビューデータなど)
- ニコニコデータセット(ニコニコ動画コメント等データ)
- リクルートデータセット(ホットペッパービューティーデータ)
- クックパッドデータセット(レシピ・データなど)
- LIFULL HOME'Sデータセット(賃貸物件データなど)
- 不満調査データセット
- Sansanデータセット(サンプル名刺データ)
- インテージ・データセット(i-sspデータ;TV、PC、スマホでのメディア接触データなど)
- その他

スタンフォード大学では、主に研究用にインターネット／Webのネットワークに関するデータをStanford Large Network Dataset Collection<sup>\*40</sup>で公開している。これらは研究目的用に供与されているのであり、商用利用は制限されていることが多い(商用利用にはこれらだけでは独自性などの点で不十分なことが多い)。

オープンデータについて触れておくと、官民データ活用推進基本法(2016年12月成立・施行)において、国及び地方公共団体はデータのオープン化への取組みが義務づけられた。これによりオープンデータ化は進んできており、社会問題の解決、新サービスへの利用による経済活性化が期待されている。オープンデータの重点8分野(電子行政、健康・医療・介護、観光、金融、農林水産、ものづくり、インフラ・防災・減災、移動)が定められ、整備が進められている。

※39 NII提供中のデータセット<<https://www.nii.ac.jp/dsc/idr/datalist.html>>

※40 Stanford Large Network Dataset Collection<<http://snap.stanford.edu/data/index.html>>

必要なデータは利用すべきである(データカタログサイト<sup>※41</sup>によると2018年7月現在で21,647データセットが公開されている)。情報処理推進機構(IPA)はオープンデータの利活用性の向上を目指して、「共通語彙基盤」を整備している<sup>※42</sup>。

以下ではビッグデータ利活用に関するいくつかの課題に関して記すことにする。

## ①プライバシー問題

ビッグデータの何割かは個人情報であり、プライバシー上で特別の配慮を要する。日本では改正個人情報保護法が2017年5月30日に施行され、これに従う必要がある<sup>※43</sup>。その要点を記すと次のようになる。

### ◇個人情報の定義の明確化

- ・「個人情報」とは氏名、生年月日、その他の記述等により個人が識別できるのに加え、以下の個人識別符号も含む。
- ・「個人識別符号」とは、身体の一部の特徴を表す符号(DNA、顔、虹彩、声紋、指紋、手指の静脈など)と、サービス利用や書類において対象毎に割り振られる符号(公的番号、旅券番号、マイナンバー、住民票コード、免許証番号、基礎年金番号、各種保険番号など)。
- ・「要配慮個人情報」とは、人種、思想、信条、政治的見解、病歴など、社会的差別を受けられる恐れのある個人情報。本人の同意を得なければ取得できない。
- ・「匿名加工情報」は、特定の個人を識別できないように個人情報を加工して得られる個人に関する情報。本人同意やオプトアウト手続きなしに第三者に提供できる。受領の第三者は、本人識別を目的として他の情報と照合してはならない。

### ◇個人情報の第三者提供

- ①目的外利用や第三者提供には本人同意が原則
- ②匿名加工情報にすれば本人同意なしで第三者提供可

#### [提供者の義務]

- ・提供先に関する記録の作成と一定期間の保存
- ・復元に繋がる情報の安全管理措置
- ・匿名加工情報に含まれる情報項目の公開

#### [受領者側]

- ・提供の記録の作成と一定期間の保存
- ・本人識別のために他の情報と照合することを禁止  
(業務の委託、事業の継承、共同利用は第三者利用に当たらない)

### ◇外国の第三者への提供制限

外国の第三者が以下に該当する場合、本人同意を得て提供できる

- ・日本と同等水準にあると認められる個人情報保護の制度を有している外国(個人情報保護委員会で定める)
- ・個人情報保護委員会規則で定める基準に適合する体制を整備している外国企業

※41 データカタログサイト<<http://www.data.go.jp/>>

※42 IPA、共通語彙基盤<<https://www.ipa.go.jp/osc/kyoutsu>>

※43 個人情報保護委員会/法令ガイドライン等<<https://www.ppc.go.jp/personal/legal/>>



医療分野に関しては研究開発の促進に資するため、次世代医療基盤法(2017年4月成立、2018年5月施行)が制定された。この法律では、改正個人情報保護法で要配慮情報に該当する医療情報も、オプトアウト方式により第三者提供を可能としている。

プライバシー、個人情報保護についての規制は各国で異なるが、EUは最も厳しい部類になる「一般データ保護規則(GDPR)」を2018年5月に施行した。GDPRの主目的は、市民と居住者が自分の個人データのコントロールの主権を取り戻すことであり、国際的ビジネスの規制を定めることである。以下にいくつかの要点を抜粋して記す。

- ・名前、識別番号、住所、メールアドレス、IPアドレス、クッキー、クレジットカードやパスポート情報といった個人データの収集には個人の明確な同意が必要。
- ・償いを済ませた後の犯罪歴といった過去のデータの消去を求める「忘れられる権利」を認める。
- ・収集された個人行動や購買の履歴等の個人データを、個人が自分の意思で異なるサービス提供者に移転できるデータ・ポータビリティの権利。
- ・プロファイリングに服さない権利(GDPRにおけるプロファイリングとは、個人の側面<業務実績、経済状態、健康、嗜好、興味、信頼、行動、所在または移動>の分析や予測をするためになされる個人データのあらゆる形態の自動的な処理)。
- ・EU規則と同じレベルの規則を課しているという十分性が認定されない限りEU市民のデータをEU域外に持ち出せない(日本政府はEUからこの十分性を認める合意を得ている)。
- ・違反した場合には巨額の制裁金(全世界のグループ売上高の4%か2000万ユーロのいずれか高い方)。

背景として、EUはGAFA(Google、Apple、Facebook、Amazon)等の米国プラットフォームに対する市場寡占に異を唱えていることも関係している。それに関連した制裁の動きもみられるようになってきており、プラットフォームの無制限に近かったデータ収集活動が今後変化する可能性もある。

個人データの扱いに関しては、以上のような法規や規則に十分配慮して対処することが不可欠である。

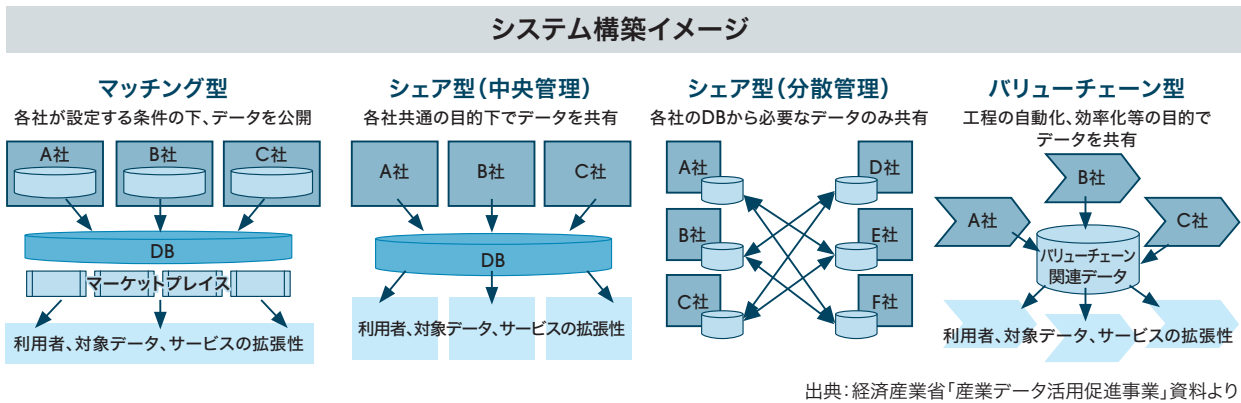
なお中国については、国民のプライバシー意識は西欧先進諸国と事情が異なっており、国家主導も含めて国民の個人データ収集と利用が進んでいる状況である。

## ②データの共有・共用の課題と、情報銀行、MyDataの動き

国際的プラットフォームGAFA(Google、Apple、Facebook、Amazon)と比べると、日本企業の1社で保有するデータ量はかなり低いレベルである。この課題を克服する一つの有効な方策は、複数企業でデータの共有・共用を可能にすることである。

この考えの下、経産省では2018年に「産業データ活用促進事業」を開始した。データ共有・共用を可能にする基盤構築と実用的運用、データ標準化・互換性、API連携の実現を狙いとしている。図2-4-2はこのような複数企業によるデータ共有・共用のいくつかの形態を示している。公募により企業から提案を受け、事業が開始されている。

■ 図2-4-2 複数企業によるデータ共有・共用のいくつかの形態



民間では、一般社団法人データ流通推進協議会<sup>\*44</sup>で取組みが行われている。

上記は企業間のデータ共有・共用の枠組みであるが、個人から（同意を得て）個人データの預託を受け、個人が許可したデータを必要とする企業に仲介、提供する「情報銀行」も課題の検討と並行して、具体的事業も始まっている。この場合の個人データは以下のようなものが対象になる。

**購買データ、移動データ、健康データ（装着センサーデータ）、食事データ、WebやSNSアクセスデータ、金融データなど。**

データを提供した個人は対価（ポイントや特別なサービスのこともあり）を受けることができる。政府もこの動きを推進しており、データ流通、利活用促進の枠組みとして期待される。

情報銀行は「個人が自分の個人データのコントロール権をもつ」ことを前提とし、その実現が図られたときにデータの範囲は大になる。この権利はEUのGDPRの中心理念であり、具体的にはデータ・ポータビリティの権利と関わる。EUのGDPRをはじめとする「個人が自分の個人データのコントロール権をもつ」との思想は最近ではMyDataとも称されている。MyDataと称するイベントは最初2016年にヘルシンキで行われ、日本でも2017年、2018年に行われた。

このMyDataの思想に深く関係するのが Personal Data Storage (PDS) であり、個人が自分の個人データを自分で保持し管理するストレージである。情報銀行とも深く関係する個人データの管理形態であり、情報銀行は個人から同意を得たPDSの個人データの預託を受けることになる。

このように、ビッグデータの収集は各企業で進むのと並行して、企業間、個人—企業間のデータ流通の枠組みも整備が進もうとしている。

### ③データバイアス問題

人間に代わってデータから学習したデータ駆動型意思決定が行われるようになることで、収集したデータに内在する（不当な）偏りや偏見が判定に影響することが起こりうる。これは、データバイアス問題と呼ばれる。

例えば、採用審査時に応募書類をAIが（予備）審査する、あるいはローンの与信審査でまだ返済履歴がない人の場合などにAIが審査するようなことが始まっている。これらのAIはデータから機械学

※44 一般社団法人データ流通推進協議会<<https://data-trading.org/>>

習(代表的にはDNNの学習)によって得られたものであるとき、元の学習データに偏りがあるとそれを反映した判断となり、公平な判断になるとは限らないことになる。

難しいのは、偏りや偏見が社会に存在するとき、それらがデータにも含まれ、AIが学習してしまうことである。表2-4-2にデータがもたらす偏りや偏見に関する例を示す。

■表2-4-2 データがもたらす偏りや偏見の例

偏りの種類	概要	例
1)データサンプリングの偏り	特定の地域や年齢層などのデータを偏ってサンプリングすると、事実との差異が生じる。	台風の被害状況をTwitterの投稿をもとに把握しようとしたところ、投稿が多かったのはスマートフォンの所有率やTwitter利用率が多い地域であった。 <sup>※45</sup>
2)データに含まれる偏り	ネット上のデータ自体に思想的な偏りが含まれている可能性がある。	米国の調査では、45%の回答者がメディアに政治的バイアスがあると回答(過去最高) <sup>※46</sup>
3)意図的な偏り	開発者や学習させる者が、意図的に偏見をもたせる学習を行う可能性がある。	Microsoftが公開したユーザーとの会話で学習するAIが不適切な発言を連発するようになり、公開停止に。
4)統計的データがもたらす偏見	統計的データが受け取り手に偏見をもたらし可能性がある。	ニューヨーク市警察が公表しているレポートでは殺人事件の被害者も容疑者も黒人が50%を超えている。 <sup>※47</sup>

特にDNNの判断結果の根拠の説明を得ることが(現状では)難しいので、公平な判断結果か否かを検証することが難しい。そもそも上表の2)の例では、回答者である個々の国民の思想にも偏りがありうるため、何が中立かの定義すら難しい。また、4)の犯罪レポートのように、正確な統計データを学習させるとAIに偏見を持たせるように見え、学習させなければ分析に支障をきたしうるようなケースもある。

このデータバイアス問題に対しては、できるだけ偏りや偏見のない学習データを用意すること、学習内容をユーザーに開示すること、AIが明確な偏りを持っていないかをチェックする仕組みを設けることくらいしか、現状では良い解決策は見出されていない。

#### ④ 学習に大量データを要する課題

DNNの学習には大量の訓練データを要することが、実用上の課題になることが多い(教師付き学習用には判定ラベル<教師データ>をもつ訓練データを必要とするが、ラベル付けを人手で行わなければならない場合は大変である)。このデータ量の課題への対処法を簡単に記す。

データ拡張(Data Augmentation)は、存在するデータに変形、変換を加えて、データを水増しする方法である。画像データの場合は、以下のような変形、変換した画像を生成して、訓練データに加える。

左右・上下の反転、回転、移動、拡大・縮小、ノイズ付加、コントラスト調整、ガンマ変換、色変化

※45 Harvard Business Review “The Hidden Bias in Big Data”<<https://hbr.org/2013/04/the-hidden-biases-in-big-data>>

※46 A GALLUP/KNIGHT FOUNDATION SURVEY “AMERICAN” VIEWS: TRUST, MEDIA AND DEMOCRACY <<https://knightfoundation.org/reports/american-views-trust-media-and-democracy>>

※47 New York City Police Department “Crime and Enforcement Activity Reports” <<https://www1.nyc.gov/site/nypd/stats/reports-analysis/crime-enf.page>>

データ量が少ない場合には過学習が生じやすくなるが、データ拡張は過学習対策としても効果を持つ。データ拡張は画像認識においては効果が確認されており、標準的な方法になりつつある。

DNNの転移学習(Transfer Learning)は、大量訓練データで学習済みのDNNを利用し、類似性はあるが別の新しい認識対象に対して学習済みのDNNの最終層近くの数層のニューラルネットワーク部分のみを微調整(fine tuning)する学習を行う。この場合、微調整部分以外の部分(低レベル、中レベルの特徴部分に相当)は学習済みDNNと同じパラメーターをそのまま使い、最終層近くのニューラルネットワーク部分は比較的少量の新たな対象物データで微調整(学習)することにより、新たな対象の認識/識別を行うDNNを得ることができる。

この転移学習も画像認識/識別ではよく利用されるようになってきている。VGG16(オックスフォード大で開発された16層から成る畳み込みDNNで、2014年のILSVRC<ImageNet Large Scale Visual Recognition Challenge> 画像認識で2位であった)は、転移学習用の学習済みDNNとしてよく利用されている。

人間は過去に学んだ何らかの関連知識を利用している面はあるものの、画像認識/識別における対象物の学習にDNNのような大量データはなしでも行っている。この点に着目し、少数データで可能な機械学習法(DNNには限定せず)は、日本も含め世界で活発に探究されている。

### 2.4.3 知識を巡る状況

ビッグデータ時代において、AIで価値を生み出すソースとしてデータの注目度は高まっているのに対し、(記号的)知識への注目度は相対的に下がっている感もあるが、前述のように知能の思考/推論の側面を支える知識は重要であり、着実な歩みを見せていると言える。WebやSNSではテキスト部分は非構造化データでもあるが、テキストマイニングにより知識を抽出して利活用が図られるので、データと知識の中間的な位置になると言えよう。ここではこのようなものも含めて、知識に関する状況を記す。

Webは情報共有、流通のプラットフォームとして登場し、1990年代半ばからのその普及と拡大は検索エンジンを伴うことで、情報環境に革命的变化をもたらした。2000年代になると、BlogやSNSも登場し、いわゆるCGM(Consumer Generated Media)によって、各個人のバラエティに富む情報発信も増大した。

知識の点で注目すべきは2001年に始まった共同執筆のWebオンライン百科事典であるWikipediaである。2018年7月時点で英語版は記事数で567万、日本語版で記事数111万となっている。内容は中立的観点から記述すべきとの方針がとられており、情報の質も2005年の百科事典エンサイクロペディア・ブリタニカとの比較で大差はなかったという調査結果も示された(その一方で、査読制度がなく、問題ある記述はコミュニティでの管理に委ねられているので、信頼性の担保が弱い部分もあるとの意見もある)。Wikipedia情報をコンピューターによる意味把握を容易にするLOD(Linked Open Data)形式にしたのがDBpedia<sup>\*48</sup>である(LODについては後述する)。

Webの発明者であるTim Berners-Leeは1999年にセマンティックWebを提唱した<sup>\*49</sup>。通常のWebは人が読む文書情報をHTML形式で表すのに対し、セマンティックWebはコンピューター(機械)が意味内容を判読できるようにすることを目標とする次世代Webの構想である。セマンテ

---

\*48 <<https://wiki.dbpedia.org/>>

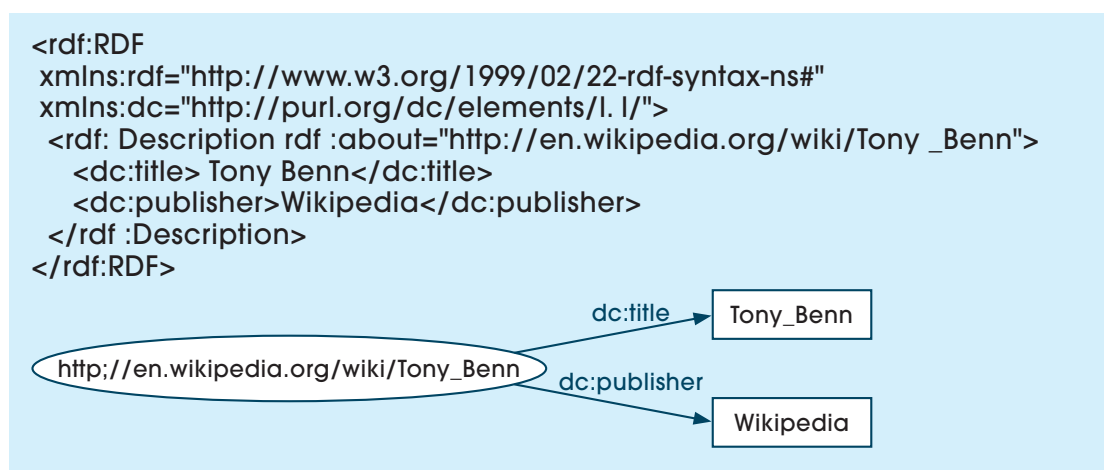
\*49 Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor., Tim Berners-Lee, with Mark Fischetti. Harper San Francisco, 1999.

ィックWebの上位目標(Proof、Trustなどあり)は高く、研究開発や標準化は継続しているものの、なかなか普及するまでに至っていない。

セマンティックWebの基礎となるデータ記述形式はRDF (Resource Description Language) であり、これにもとづいてコンピューターに意味が判読でき、Webのようにオープンにアクセスできるようにしたのが、2007年からLODと呼ばれるようになったデータのWebである。データであるものの、概念や事物間を豊富な関係で結んでおり、知識化されているとも言える。

記述形式のRDFは主語(subject)、述語(predicate)、目的語(object) の三つ組(triplet) でデータ/知識を表す。図2-4-3にRDFの例を示すが、下がグラフ表示であり、上がNotation3 (N3) というXMLをベースとするRDFのシリアル化記述形である(他のシリアル化記述形も存在する)。

■ 図2-4-3 RDFの例—下がグラフ表示で上がNotation3 (N3) 記述



出典: Wikipedia記事「ファイル:Rdf-graph-example-TonyBenn.png」より作成

図2-4-3下のグラフの楕円ノードはURI (Uniform Resource Identifier) であり、これはWeb空間でユニークなIDとなる。上のNotation3記述の2行目、3行目はそれぞれrdf とdc (Dublin Coreという書誌情報メタデータ記述用の関係語彙セット) のXML名前空間(xmlns) が = の右側のURIで定められることを記している。4行目の <rdf:Description rdf:about=" …… " > はRDF三つ組構造データの主語(subject) 項目を記し、その下の2行で dc:title と dc:publisher (dc名前空間のtitleとpublisher) の述語関係をもつ目的語に当たるリテラル(文字列か数値) Tony\_Benn と Wikipedia に関係づけている(ここでの目的語の位置はリテラルでありURIではないが、URI を記述することもできる。また同一性を示す owl:sameAs (owlはXML名前空間) の関係述語でリテラルをURIに結び付けることができる)。このようにして概念や事物がWeb上でユニークなURIと結び付けて表されることにより、多くのLODセットが関係づけられて連携し、大きなLOD空間を形成することになる。なお、普通のWebのリンクは単に参照先のURIを示すだけであるが、RDF(あるいはLOD)のリンク(述語)は意味をもつ関係を示すという違いがある。

LODをはじめとして知識記述に必要なオントロジーにも触れておくと、オントロジーは哲学用語としては「存在論」を意味するが、AI(あるいは情報科学)では対象世界の「概念体系」あるいは「基本語彙体系」を表す<sup>※50</sup>。記述に用いる概念/語彙を定義し、それらの間の階層も含む必要な関係記述を

※50 Tom Gruber, "Ontology. Entry in the Encyclopedia of Database Systems", Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag, 2008.

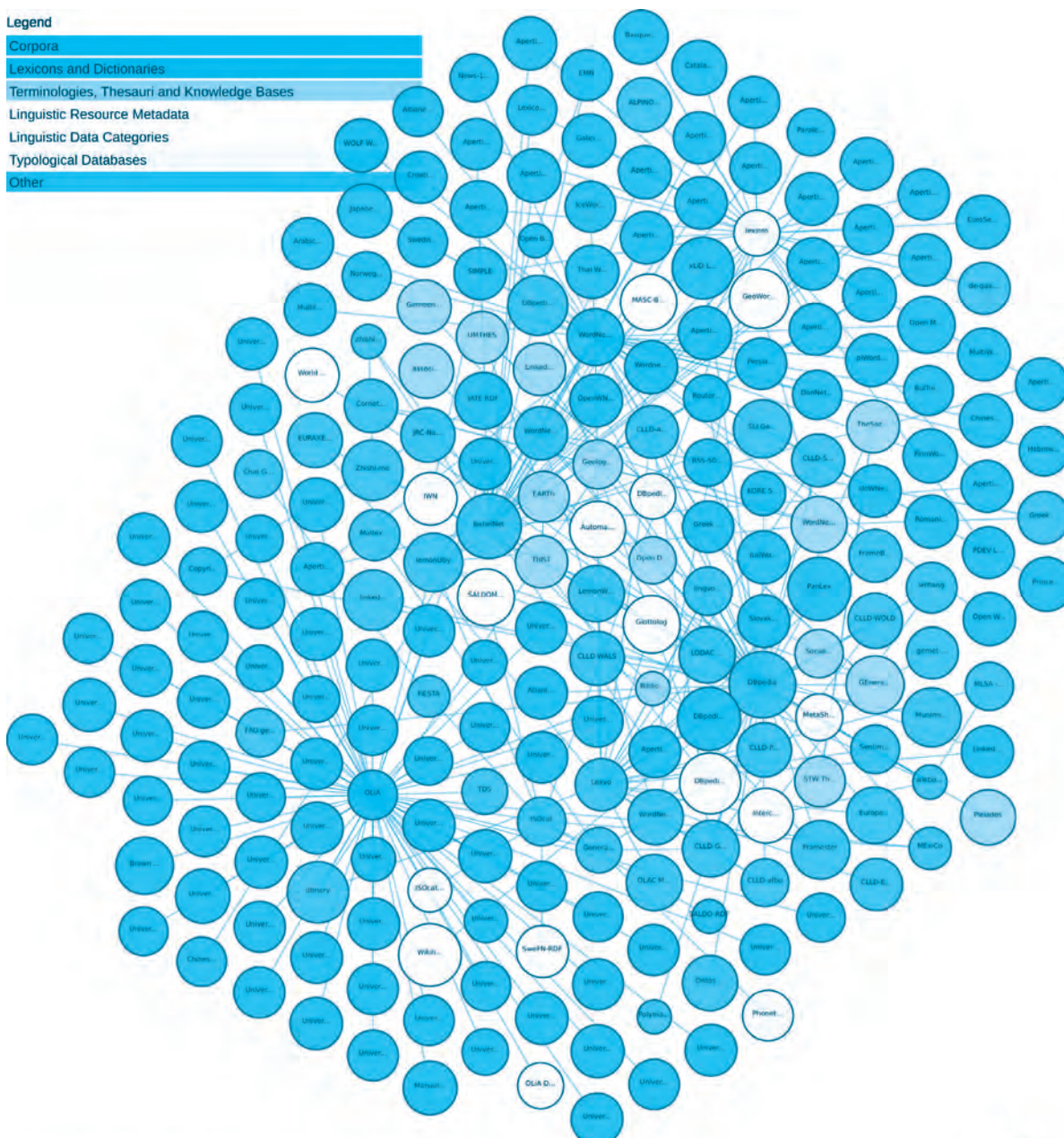
定める。ここでは深くは立ち入らないが、図2-4-3の例ではXML名前空間にはこのような体系化された基本語彙(オントロジー)が定められており、RDFの関係記述に使用されている。

LODは2018年6月時点で1,205データセットが公開されており、総体としてLinked Open Data (LOD) Cloud と呼ばれる。どのような対象領域があるかと言うと、大別すると以下のようなサブドメインである。

Cross-Domain(領域間共通)、Geography(地理)、Government(政府)、Life Science(生命科学)、Linguistics(言語学)、Media、Publication(出版)、Social Networking、User-Generated(ユーザー生成コンテンツ)

LOD全体では1,000データセット以上にもなり込み入ってしまうので、図2-4-4にはLinguisticsサブドメインのLODを例示する。各ノードがデータセットを表し、関係づけがあるノード間にリンクが張られている。

■ 図2-4-4 Linguistics(言語学)サブドメインのLOD(コーパス、用語と辞書、専門用語・シソーラス・知識ベース、言語リソースメタデータ、言語データカテゴリ、印字学データベース、その他が含まれている)



出典:The Linked Open Data Cloud「LinguisticsサブドメインのLOD」のグラフ(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/>>

代表的で最も規模が大きいLODは DBpediaであり、2018年で英語版は458万項目を含んでいる。主にWikipediaのInfobox（右上のボックス部分）の情報がRDF形式で記されている（DBpedia Japanese も存在しているが<sup>\*51</sup>、規模は大きくはなく範囲も限定されている）。

WikipediaをもとにしたLODにWikidata<sup>\*52</sup>も存在する。DBpediaがWikipediaの主にInfoboxから情報抽出しているのに対し、WikidataはWikipediaの内容に関して共同編集されたRDFを集積している（DBpediaはBerlin自由大学とLoipzig大学で始められ2007年に公開されたが、WikidataはWikimedia財団で2012年から開始された）。

WordNet<sup>\*53</sup>は（英語の）上位・下位語、同義語等を定義した語彙データセットだが、このWordNetもLODとして公開されている（日本語WordNet<sup>\*54</sup>も存在し、LOD化が図られている）。

RDFのクエリ言語として SPARQL<sup>\*55</sup>があり、RDFにより記述されたLOD空間は、SPARQL Endpoint（LODの標準検索API）からSPARQLによって検索可能である。

なおここではLODの記述形式としてRDFを記したが、RDFはオープンでないプライベートな三つ組構造で関係付けられたデータ／知識の集合を表すのにも用いられる（その場合はLinked Data と呼ばれる）。RDFという規格化された表現形と、それに付随する各種ツールも利用することが利点となる。

RDFではデータ／知識は三つ組 (triplet) 及びノードと関係リンクによるグラフで表現される。同じようにデータ／知識をグラフ表現したものであるが、RDF表記は用いない知識グラフ (Knowledge Graph) がある。

キーワード検索の検索エンジンはWeb情報空間利用に不可欠のツールであるが、雑多な情報の選別など、アクセスには不十分な点も多い。GoogleのKnowledge Graphは、Web等から抽出したオブジェクト（事物）間の関係を意味ネットワーク形（オブジェクトをノードとし、ノード間を意味的關係を付したリンクで結んだグラフ）で知識化して表したものである。これにより単なるキーワードでなく、オブジェクト間の意味的關係も考慮した検索を可能にしている。このGoogle Knowledge Graphは2012年の発表時点で5.7億件のオブジェクトと、それらオブジェクト間の180億の意味的關係を有していた。2016年には700億件のファクト関係に拡大しており、検索と共に機械翻訳の品質向上にも利用されている。これはGoogle Knowledge Graph Search API<sup>\*56</sup>から検索で利用することができる。

2016年に発表があった Microsoftの Concept Graphは、テキスト理解に必要な概念をやはりノードとし、ノード間を確率を伴う関係で結ぶグラフとしている。この概念ノード数は540万程であり、テキスト文の常識に照らした確率的解釈に役立つ。

Microsoftはこの他にもGoogle Knowledge Graphのような知識グラフ Satoriも、検索エンジンBing用や対話システム用に開発している。Satoriは2012年時点で3.5億項目、8億

※51 DBpedia Japanese<<http://ja.dbpedia.org>>

※52 Wikidata<[https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)>

※53 Word Net<<https://wordnet.princeton.edu/>>

※54 日本語Word Net<<http://compling.hss.ntu.edu.sg/wnja/>>

※55 SPARQL<<http://www.w3.org/TR/rdf-sparql-query/>>

※56 Google knowledge Graph<<https://developers.google.com/knowledge-graph>>

の関係記述をもつ。

FacebookもEntities Graphという仮称で、膨大な利用者のプロフィール情報をノードとそれら関係をリンクで結ぶグラフ形状のデータ／知識ベースで管理して活用している<sup>\*57</sup>。

データが膨大になることで、グラフ形状のデータ／知識ベースの処理の高速化も、解決すべき課題になっており、米国防高等研究計画局(DARPA)は、「Hierarchical Identify Verify Exploit ; HIVE」と呼ばれる高速化の研究に取り組んでいる<sup>\*58</sup>。

1980年代半ばから常識(common sense knowledge)の知識ベース化の先駆的活動であり、主として人手に頼って構築が進められたCYC<sup>\*59</sup>は、上記のようなテキスト・ビッグデータからオブジェクトや概念を自動抽出している知識グラフに比べると、(質はともかくとして)量の面では対抗できないようである。

WebやSNSとは別の知識源として重要になってきているのが科学論文であり、テキストマイニングにより知識を抽出し、活用が図られている。代表的な例としては医学／医療論文データベース(代表としてMEDLINE—PubMedを介してアクセスされる)は、医療や創薬等によく活用されている。背景には多数公表される科学論文は専門家でもとてもフォローし読み取ることができない量になってきている状況がある。

類似な例としては特許文書などもある。企業での文書や資料もテキストマイニング等により加工、整理され、知識化しての利用が進むことが期待されている。

知識の使い方として近年目立ってきているのは質問—応答(Q&A)用の知識である。コールセンターでの対話応答(オペレーターの応答支援と自動応答あり)、スマートスピーカーでの対話、チャットボットでの対話などを、正しくかつ円滑に行うための知識の蓄積(知識ベース)である。企業内での様々な知識(ファクト型知識とknow-how的あるいは手続き型知識あり)の共有と流通(問合せに回答する等の形式)も近い形態といえる。

第2次AIブーム期の知識ベースはルールを主体にして、一部は構造をもつフレーム、論理、制約等の形態で知識が集積されたが、今日ではルールは用いられるものの(高速化のReteアルゴリズム<sup>\*60</sup>を有する)プロダクション・システムの使用などはあまり聞かなくなっている。制約は知識の基本的なものであり、随所に使われ、充足可能解あるいは最適解を導出するアルゴリズム(一種の推論)が使われる。フレーム名のオブジェクトが関係をもつ要素をスロットして表すフレームの機能は、前述した三つ組(triplet)表現をとるRDFやその類似形で代替されてきている感がある。

データ利活用の観点からは、オープンデータは誰でも自由に活用できることがその骨子であり、技術の民主化、すなわちLODとその活用技術の普及をボトムアップに応援する活動も多く存在する。代表的なものはLODチャレンジJapan実行委員会が主催する、「LODチャレンジ」であり、コンテスト形式で評価し合うイベントを通じて、LODの技術情報を発信するとともに、データやアイデアに関する情報交換や共有を行うコミュニティづくりが行われている<sup>\*61</sup>。

---

\*57 Facebook Entities Graph<<http://www.facebook.com/notes/facebook-engineering/under-the-hood-the-entities-graph/10151490531588920>>

\*58 HIVE<<https://www.darpa.mil/program/hierarchical-identify-verify-exploit>>

\*59 Cycorp/Cyc<<http://www.cyc.com>>

\*60 Rete Algorithm<[https://en.wikipedia.org/wiki/Rete\\_algorithm](https://en.wikipedia.org/wiki/Rete_algorithm)>



一方、データの重要性が注目されている中でいわゆる知識ベースの形態や実体では、LODなど標準化されオープン化されているものは把握できるが、企業内のもの等は様々な形態であり、全体像を把握しにくくなってきている。

先にも述べているように、深層学習(DL)等の機械学習によって得た、認識/識別の用途に代表的に使われる学習済みDNN等是一種の知識であり、これらの獲得した知識の流通・共有の枠組みは、データの流通・利用と並行して今後の課題である。学習済みDNNについては、訓練データとは違うが類似性のある新たな認識対象に対して、転移学習による利用が可能ではある。しかし、実際に転移学習に利用されるのはImageNet で学習されたVGG16を代表とした限られた範囲である。学習済みDNNをはじめとするデータから学習によって得た知識の広範囲な利用を可能にする枠組みは、適用範囲や品質の保証をどうするかなど、今後の課題である。

---

※61 <<http://2018.lodc.jp/>><<http://2017.lodc.jp/>>等々

## 2.5 ▷ 身体性とロボティクス

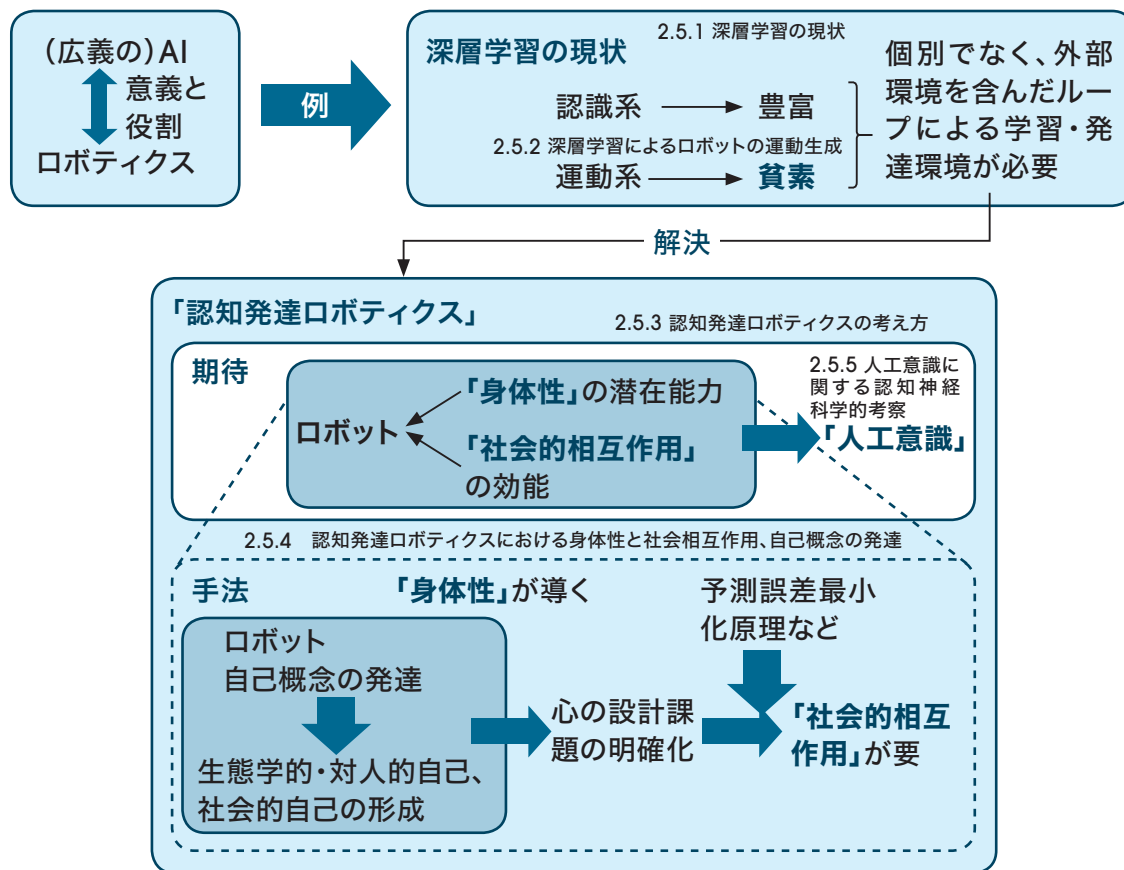
本節では、近年、目覚ましい進展を遂げ、分野を拡張しつつある広い意味でAIの観点からロボティクスの役割と意義の説明を試みる<sup>\*62</sup>(図2-5-1)。この観点から、最初に現状の深層学習の状況を再認し(2.5.1)、認識系のデータ量、結果、応用範囲の豊富さに比して、運動系が量的にも質的にも貧素であることを指摘する。そして、それを打破する可能性を示す例を紹介し、残された課題等を列挙する(2.5.2)。

これを解決する一つとして、認知発達ロボティクスの考え方を導入する(2.5.3)。次に認知発達ロボティクスの思想的背景である人間の心と身体や事物との関係に関する歴史的流れを概説する。それは、ロボットを持つ身体性や社会的相互作用の能力が、AIの一つの究極的課題である人工意識に結実されるのではないかという期待からである。それにより心の設計課題が浮き彫りになると予想される。

そして、認知発達ロボティクスのゴールである人間の認知発達過程の理解の一つの側面として、ロボットが自己という概念を持つ可能性を検討する。自己概念の発達の形態として、生態学的自己、対人的自己、社会的自己を形成する機構を検討する(2.5.4)。

それぞれの段階での課題を列挙し、ロボットの身体性にもとづく、身体表象、自己感覚、共感などの社会的相互作用に関連する認知発達ロボティクスのアプローチを紹介する。前半は計算神経科学的側面が中心であり、後半は、認知発達過程の計算原理としての予測学習規範を紹介し、その可能性を検討する。そして、最後に人工意識の在り方を議論する(2.5.5)。

■図2-5-1 本節の全体像

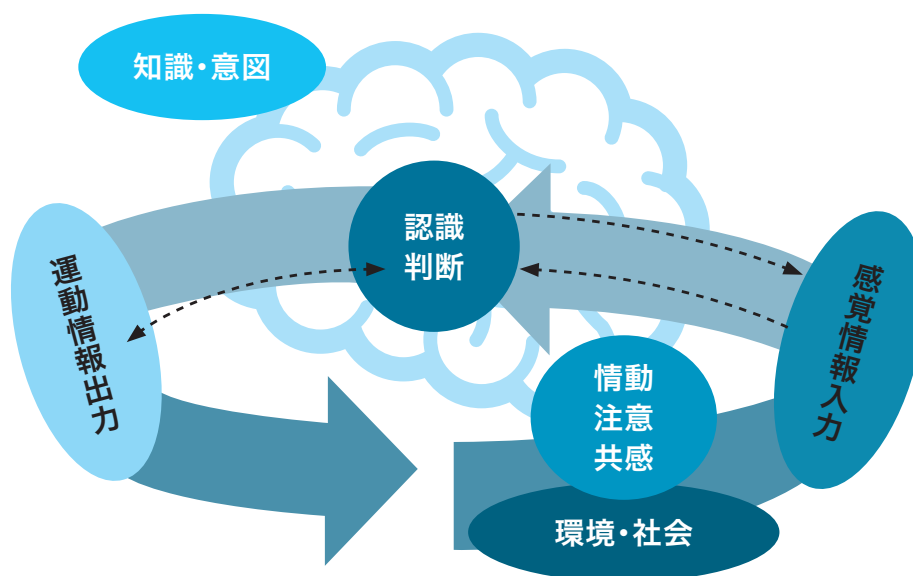


※62 ロボティクス分野一般の在り方に関しては、やや古いですが、別の成書[1]を参照していただきたい。

### 2.5.1 深層学習の現状

深層学習の現状の適用範囲を図2-5-2に示す。随時、更新改良がなされているので、すぐに古くなることを覚悟で示す。深層学習は、もともとヒトの視覚システムのモデル化から始まっているので、音声も含めたマルチモーダルな感覚情報からの認識・判断が得意であり、言語データとも組み合わせた音声や画像の出力も可能である。片や運動出力に関しては、その厚みがない。理由は、画像の場合、数千万枚のデータが利用可能であるが、ロボットの試行回数には限界があるからである。運動出力とのパイプが太くなることで初めて、環境を含んだループが完成し、入力(感覚)から出力(運動)への即応的な応答(自動的かつ無意識的な行動)が可能になる。最近、やっとロボットの行動生成への適用が出始めている。これらに関しては、次項で触れる。

■ 図2-5-2 深層学習の現状の適用範囲



### 2.5.2 深層学習によるロボットの運動生成

「2.2 ディープラーニング」で触れた深層学習では、主に多様な感覚情報からの認識・判断の学習がメインであり、優れた成果が出ていることから、これをロボットの運動生成に応用する際に、以下の2つの考え方がある。

- ①これまで困難と想定されていた視覚情報処理の部分に適用し、正確な物体認識や分割領域情報を用いて、従来のロボティクスの枠組みで解決する場合
- ②図2-5-2に示したように、ロボット学習の枠組み、すなわち感覚・運動ループの中で深層学習の手法を適用する場合

前者の例としては、Lenzら [2]は、深層学習により対象物の距離画像から把持ベクトル(ロボットハンドの位置と方向)を出力した。また、Redmon and Angelova[3]は、RGB-D画像から、畳み込みニューラルネットワークを利用し把持ベクトルを予測した。さらに、Pinto and Gupta[4]は、700時間ロボットを駆使して5万回の試行データにもとづき把持の学習を行った。これはロボットの把持を対象としてはいるものの、物体画像から把持ベクトル、把持領域などのマッピングのみを問題としており、純粋なコンピュータビジョン研究で、実際の動作自体は、把持ベクトル

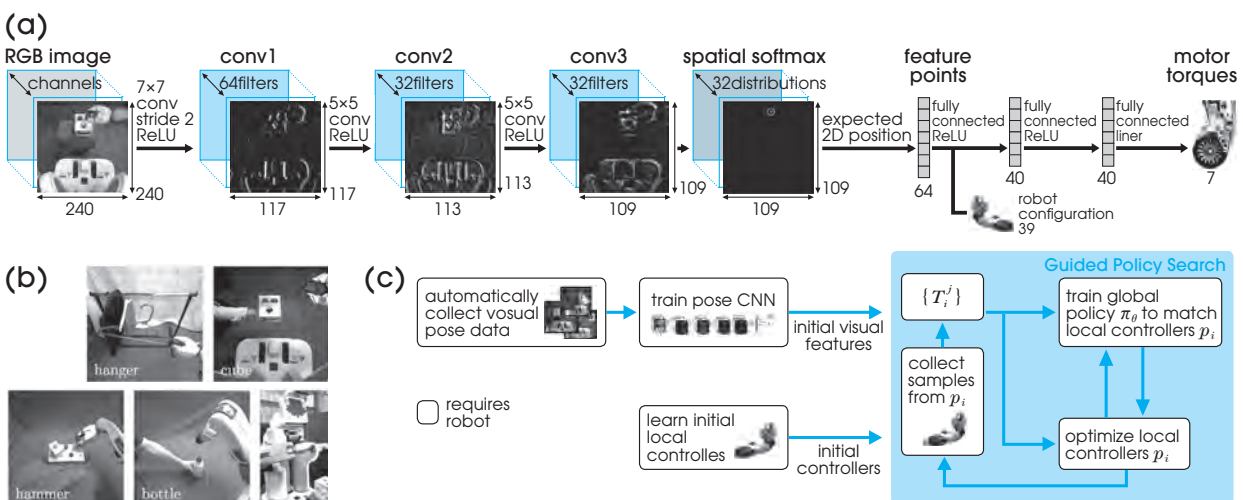
ルの情報を受けて逆運動学などの従来のロボット制御で実行されることを前提としていた。把持ベクトルが画像から得られたとしても、実際にどのような動作をすることで把持可能なかを学習対象にするには、動作を生成する身体構造も含まれる必要があるが、それらは対象となっていない。このように、単に視覚モジュールを取り替えたにすぎず、ロボットの行動制御は伝統的な手法に依存している。これは、既存手法が、入力から出力まで概念的に複数の段階の処理が必要な場合には、個々の処理をステップバイステップで学習した後にそれらを統合するという手順を踏襲しているからで、深層学習を便利で有用な視覚モジュールとみなしている。

これに対し、もともと深層学習は、入力から出力までを一つのネットワークとして表現し、全体を“End to End Learning”する形態をとって、ロボットの動作を学習する機構で、入手可能な高次の入力データ（画像や映像）から、必要な高次の出力（複数の関節時系列出力）を直接得るパラダイムである。ただし、先にも述べたように、画像認識の場合、数千万枚の画像データが利用可能であるが、実際のロボットには、そこまでの試行回数を求められない。なんらかの工夫やバイアスが必要である。

### (1) 視覚運動方策のエンド・ツー・エンド学習

Levineら[5]は、PR2と呼ばれる双腕ロボットプラットフォーム(図2-5-3(b)右下)に深層強化学習による行動探索を行わせ、現時刻の1枚の視野画像入力から次時刻のロボットの複数関節出力(7軸のトルク指令)を直接CNN(畳み込みニューラルネットワーク)で出力(図2-5-3(a))させ、一連の動作を実現する手法を提案し、複数の動作について、物体位置変化などがあっても安定的に動作させることに成功した(図2-5-3(b))。ポイントは図2-5-3(c)に示すアプローチで、画像から姿勢を決めるCNNと局所コントローラーを事前にトレーニングし、それらを用いて、Guided Policy Searchと呼ばれる方法で、大局的な方策を求めることである。初期トレーニングによるバイアスと探索手法に工夫がある。画像特徴は、単なる画像上の特徴点(エッジや孤立点など)ではなく、タスクを遂行するうえで必要な特徴点を抽出可能であり、それはロボットの身体に拘束された注意点でもある。目標点は画像上の座標で指定され、視覚サーボによりロボットを制御するが、そのためのカメラキャリブレーションは、CNNの学習に内包されている。ちょうど、構造やパラメーターに関する事前知識を必要としない適応型ビジュアルサーボシステム[6]のように振る舞う。

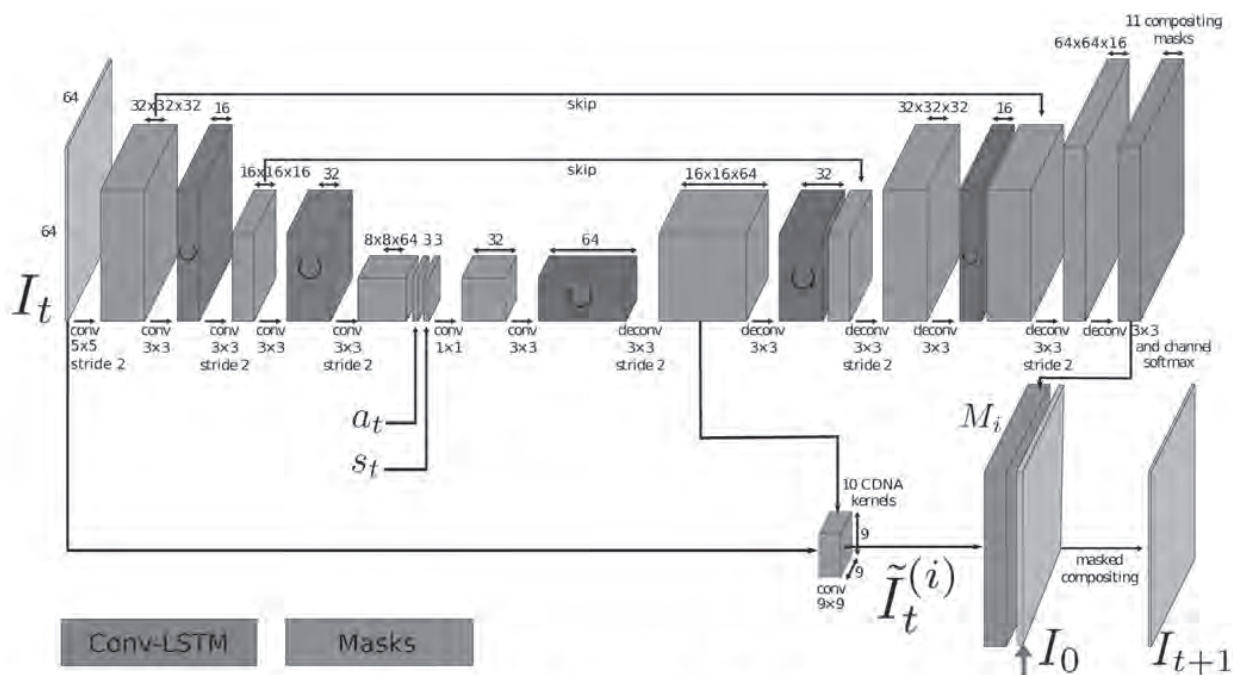
■ 図2-5-3 深層学習による視覚運動方策のエンド・ツー・エンド学習: (a) 視覚運動方策アーキテクチャー、(b) 様々なタスクへの応用 (ハンガー掛け、ブロックの挿入、ハンマーによる釘抜き、キャップの締め) とロボットプラットフォーム PR2、(c) 探索アプローチ



## (2) 長短期記憶構造と報酬関数を組み込んだ深層学習による時系列画像の予測と行動生成

(1)で紹介した視覚運動方策学習では、現在の画像から現在の状態を推定し、必要なモータコマンドを生成していたが、様々な応用や3次元空間での隠蔽などの状況に対応するために、同じ研究者グループは、7自由度ロボットアーム10台による5万回のプッシングの試行のビデオデータをもとに、時系列ビデオ画像の予測による行動生成(プッシングのみ)を実現した[7]。図2-5-4にそのアーキテクチャーを示す。

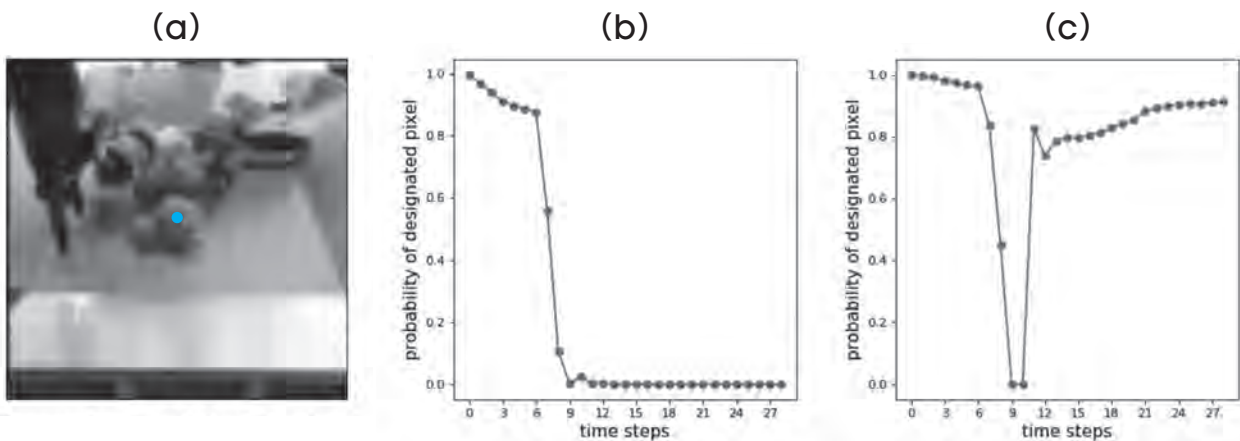
■ 図2-5-4 時系列ビデオ画像の予測による行動生成のアーキテクチャー



出典: 論文[7]より引用

時系列ビデオ画像の予測自体は、Taniらの研究(書籍[8]の8、9章辺り)でも行われており、彼らの先行研究でも達成している[9][10]。2.2節で示したように、RNN(リカレントニューラルネットワーク)は時系列信号の記憶と予測に使われるが、その能力を高めたLSTMと呼ばれる長短期記憶構造を有するニューラルネットワークを多層利用している。画素ごとの予測であり、物体の切り出しなどは一切ない。先行研究では、隠れが生じない場合は、ほぼ正しく予測できたが、ロボットアームによる隠蔽が一旦生じると、隠蔽がなくなっても、それ以降の画素の予測が正しくできなかった(図2-5-5(b))。それに対し、ロボットアームの動きによる画素の明度変化が予測される際に、時系列予測の対象区間をスキップする構造を入れると、隠蔽がなくなり、隠された物体が現れたときに正しく画素の明度を予測できた(図2-5-5(c))。ポイントは、マスキング画像( $M_i$ )で画素ごとの出現確率を保持し、隠蔽の現象の有無を判断し、最初の入力画像( $I_0$ )とマスク画像の積で次の時点の画像( $I_{t+1}$ )を予測している点である。当然のことながら、予測の時間窓の幅( $\tau$ )により判断の精度が異なる。まさに「目に入らないものは、意識されない」Out of sight, out of mindである。

■ 図2-5-5 隠蔽の予測



(a) 対象画素 (青の点)、(b) 隠蔽後、画素値が正しくない、(c) 隠蔽後、ほぼ元の画素値に戻っている

出典:論文[7]より引用

発達的な視点で考えると、記憶幅の容量が少ない、すなわち、ワーキングメモリが働かないと、隠蔽の判断ができず、見えないものは存在しない羽目になる。記憶幅の容量が増えると、隠蔽が正しく判断でき、結果として、物体の永続性の概念が獲得されることになる。画素ごとの予測という徹底した画像のみによる処理で、通常の特徴抽出や物体の切り出しはないので、画像から画像の end-to-end の学習法である。ロボットの状態は 3 次元の位置と姿勢で 6次元であるが、ロール軸は固定されているので、残りの 5次元ベクトルで表現され、現在と目標の 2つのベクトル ( $s_t, a_t$ ) が画像系列と一緒に入力されて、ロボットの状態と画像との関係が学習される (キャリブレーション過程とみなせる)。これにより、運動指令は画像内の座標として与えられ、視覚サーボにより目標に移動する。さらに、ゴール到達はゴール地点への画像上の移動のコスト関数で定義され、障害物回避は、障害物がそこに留まる (プッシュされない) ことをゴールとする別のコスト関数を設定し、マルチタスクとすることで、結果として自動的に障害物回避の軌跡が生成された。今回はプッシングという単純な作業で画像上での画素の単純な移動で表現できたが、今後、様々なタスクを想定した場合、タスクごとのデータセットによる学習に加え、それらを取りまとめるメタレベルの行動の構造化学習が必要である。

### (3) 直接教示バイアスによる折り畳みタスクの深層学習

乳幼児の発達の視点で考えると、先に示したワーキングメモリの容量という物理的な制約に加え、社会的環境の中で養育者をはじめとする他者が明示的/非明示的に乳幼児に対して教示していると考えられる。乳幼児からすれば、模倣学習である。なぜ模倣が可能かに関しては、以降の項でより詳細に議論するので、これ以上、ここでは踏み込まないが、いずれにしろ、この種のバイアスが、無限の試行を回避し、運動学習の効率を上げているのは確かであろう。

### 2.5.3 認知発達ロボティクスの考え方

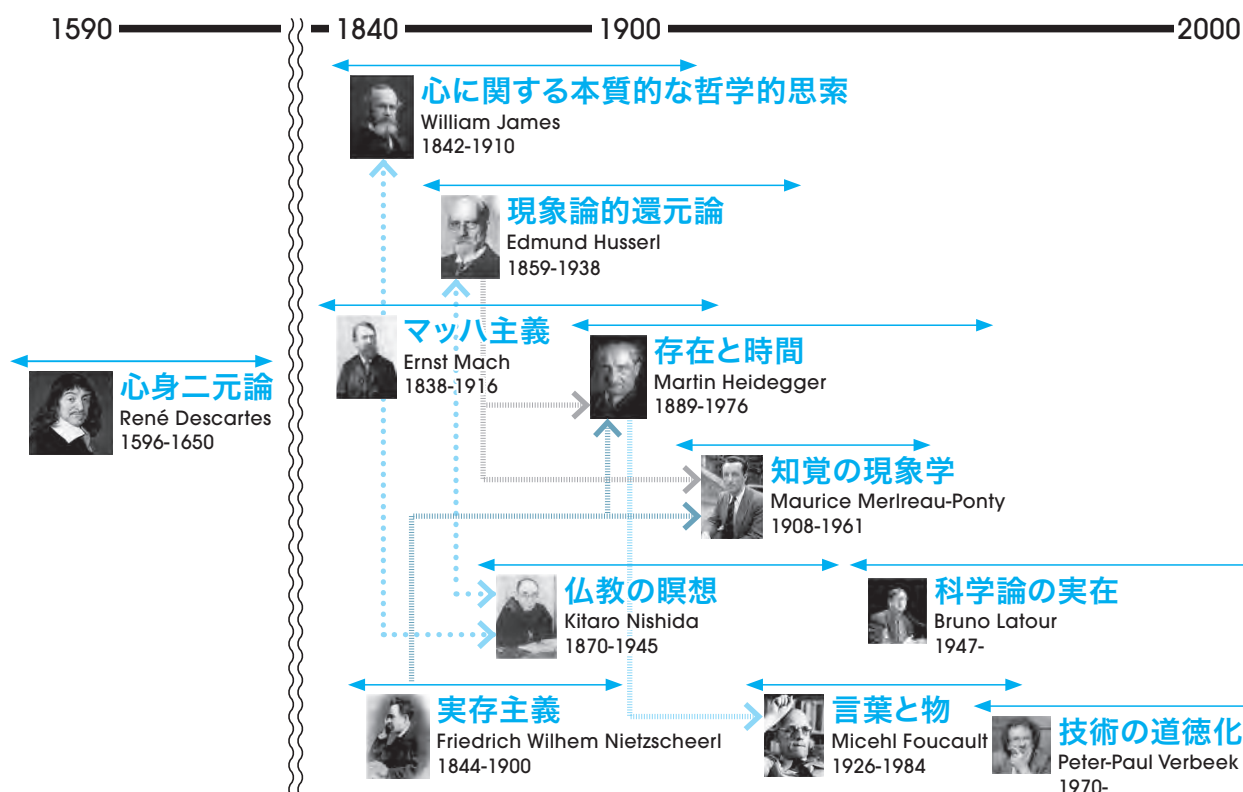
「2.5.1 深層学習の現状」で示したように、感覚運動ループによる学習は、環境内での行動学習の基本であると同時に、知識・意図の拘束条件が加わることで、専門家のスキル（経験知による行動規範）として獲得される。また、身体が環境にさらされることで、限られたリソースの下で、実時間の応答の要求に応じるために、関連しないものを処理対象としない注意が必要になり、さらに、社会的な環境の中での他者とのやり取りを通じ、情動、そして共感などのメンタルな機能が人工システムにも芽生える可能性がある。この課題を追求してきたのが、認知発達ロボティクスである。その目的は、ロボットや計算モデルによるシミュレーションを駆使して、人間の認知発達過程の構成的な理解と、その理解にもとづき人間と共生するロボットの設計論を確立することである。その中核となるのは、ロボットの身体性とそれにもとづく社会的相互作用である。その発展形が構成的発達科学であるが、目指すところはほぼ同じである。基本的な神経構造から始まり、身体性や社会的相互作用にもとづき、学習手法を介して、機能分化が段階的に生じる過程を描いている（詳細は、『AI白書2017』を参照）。以下では、まず、その思想的背景から見ていこう。

#### (1) 人工物との共生社会への思想的背景

ここでは、ものと人間の関係の哲学的視点から、もの（ロボット）自身が心のような機能が実現可能かを論じ、その過程の中で身体性の意味を再考する。

心の課題を考察にするに当たって、すべての思想的背景を網羅的に探ることは不可能に近く、以降に紹介する事項に絞って、意識・人間や事物（技術）の課題の思想的背景の概略を示す。前半はTaniの書籍[8]で紹介されている流れを、後半は、稲谷の文献[11]にもとづきながら紹介していく。

■ 図2-5-6 意識・人間や事物（技術）の課題の思想的背景の概略



心と身体、もしくは事物の関係に関して、心身二元論<sup>\*63</sup>を唱え、近代哲学の基礎を築いたのはデカルトであろう。その後、多くの批判にさらされ、いくつかのバリエーションも存在しているが、後に述べる法制度では、根強く生き残っているようである(図2-5-6)。

デカルトを超えて超越論的現象学へと進む「新デカルト主義」を主張し、現象論的考察を与えたのは、フッサールである(例えば、[12]など)。主観と客観の狭間の間主観性の考え方を展開し、後世に多大な影響を与えた。自然界の解析は、個人の意識経験にもとづくと言く。フッサールの現象論を拡張・進化させたのが、ハイデガー(『存在と時間』[13][14])やメルロ＝ポンティ(『知覚の現象学』[15][16])である。

ハイデガーは、主観と客観を分けずに実存を問い、「現存在」は、個々のエージェントの将来の可能性とその過去の可能性との間の動的相互作用によって生まれていると主張する。また、それぞれの個々が目的をもっていかに相互作用しているかという事前の理解のもとに個々が相互に存在しようという、ある種の社会的相互作用の重要性を指摘している。

メルロ＝ポンティは、主観と客観に加えて身体性という次元が創発し、そこでは、同じ肉厚の身体が、触れたり見たりする主体と同時に触れられたり、見られたりする物体にも与えられうるとし、主観と客観の2つの極の間の繰り返される交流の場を身体が与えると主張する。すなわち、客観的物理世界と主観的経験を結ぶメディアとしての身体の重要性を指摘している。これは、後に紹介する認知発達ロボティクスにおける「身体性」の基本概念の根幹である。

ラトゥール[17]は、主体と客体とを厳格に区分する近代的な思考法(人間存在の在り方を本質化するヒューマニズム)のせいで、我々は、我々が現実生活しているところの主客が入り混じったハイブリッドな世界を適切に取り扱うことができなくなっていると警告する。また、フェルベーク[18]は、「技術は、我々の行為や世界経験を形成し、そうすることによって、我々の生活の仕方に能動的に関わっている」と主張する。彼らの考え方は、高度に発展した人工システムとの関わりが、これまでとは異なるステージに入っていることを指摘し、そのための規範づくりの重要性を指摘している。AIやロボットはその急先鋒であり、研究者自身が技術的側面だけではなく、社会的な意味や価値を考慮して設計しなければならないことと、受け入れる一般大衆もそれに対する準備が必要であると指摘する。これについては、以降では触れない。解説[19]などを参照されたい。

## 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

### (1) 自己と他者の概念を確立する発達過程

認知発達ロボティクスの基本構成概念である身体性や社会的相互作用に関しては、『AI白書2017』を参照していただくとして、ここでは、メルロ＝ポンティの身体性概念[20]を発展させていくことで、各種身体表現や自己の概念などが獲得され、ミラーニューロンシステムを通じて、情動感染から情動的共感、認知的共感に至る過程において、社会的相互作用が見出されることを示していく。これと並行して、自他認知の発達もみられる。

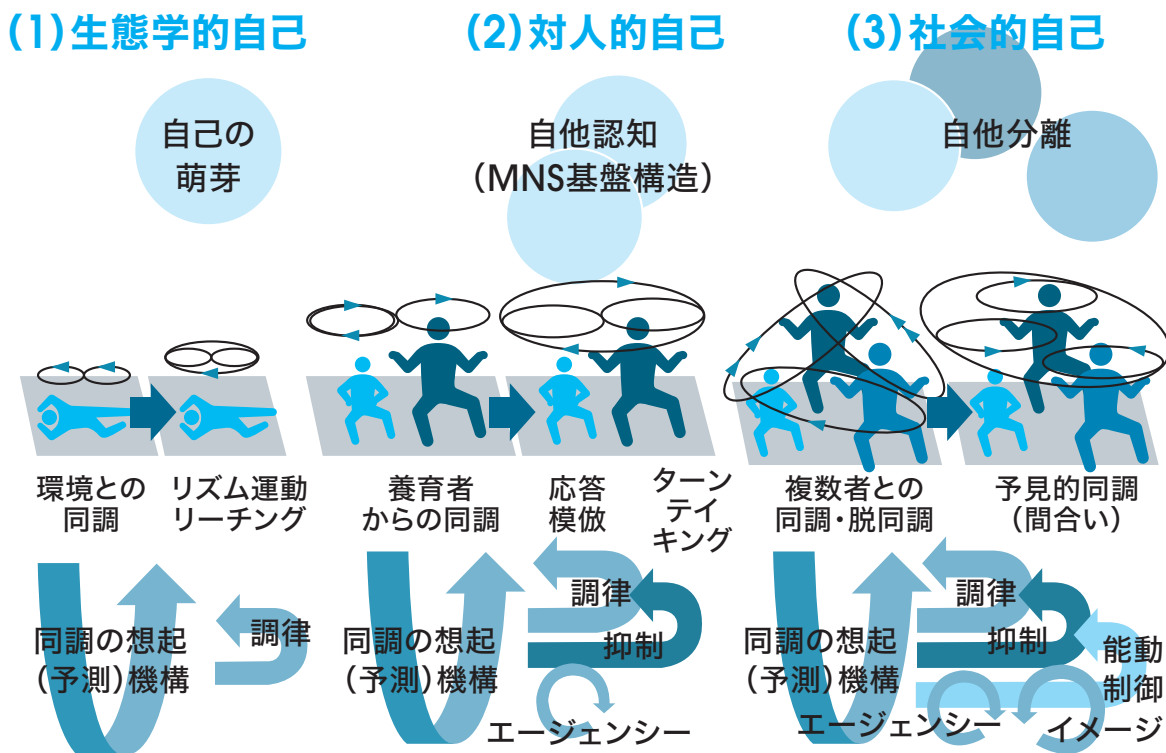
図2-5-7の上段は、Neisserの自己の定義[22]を部分的に利用した自己と他者の概念形成の発達過程を表している。同調という用語は、他者を含む外界との相互作用を通じて、この概念がどのように発達したかを説明するキーワードである。図2-5-7に示しているように、同調のターゲットは、

\*63 実体二元論、物心二元論、霊肉二元論、古典的の二元論などともいわれているようである。<<https://ja.wikipedia.org/wiki/実体二元論>>



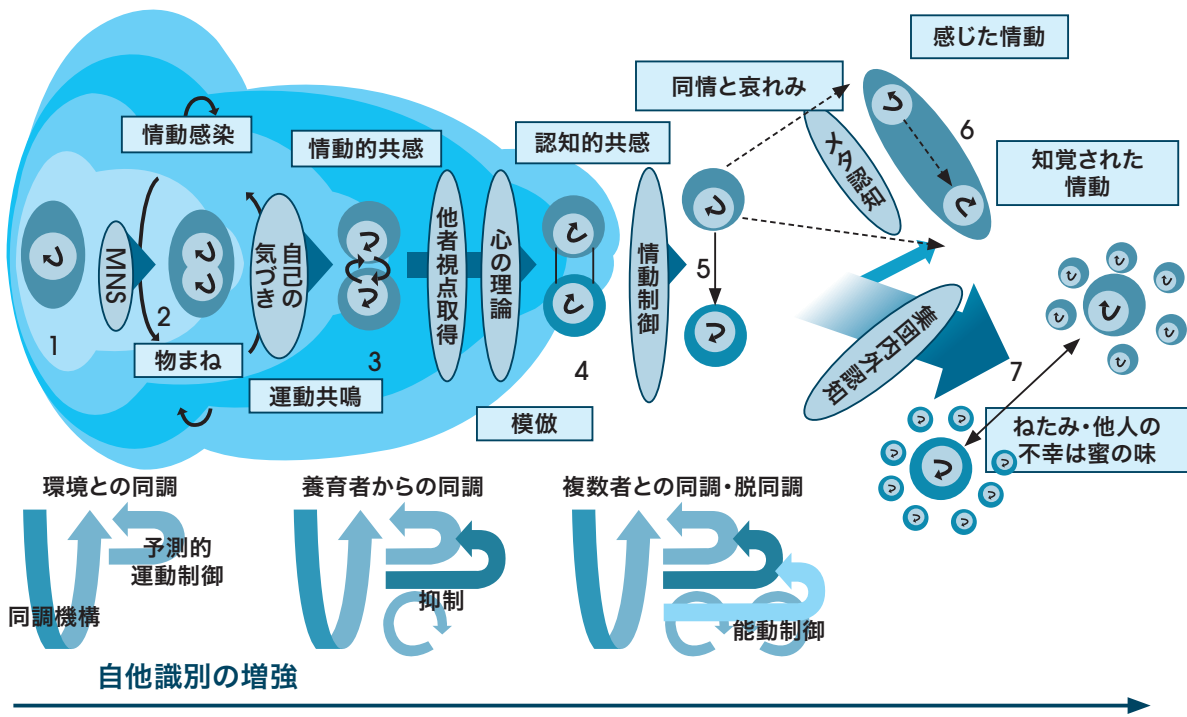
物理的物体から始まり、他者の動き、そして最終的に他者の心的状態と変化する。従って、行動も初期のリズミックな運動のプリミティブから、それらが構造化されたもの、さらには、心的状態によって引き起こされる共感的／同情的な顔表情や利他的行動にまで至る。自己の発達三段階は、実際はシームレスにつながっていると想定される（詳細な議論は、文献[23]を参照）。

■ 図2-5-7 自己と他者の概念を確立する発達過程（上）とそこに期待される機構（下）



自他認知過程は、共感の発達（進化）とも強く関連する。共感の発達（進化）過程をまとめると図2-5-8になる。左から右への流れは、自他認知と情動制御の発達と進化の方向を示している。小円の中のカーブ矢印は、エージェントの内部状態を表し、自他を識別できない状態(1)から、完全に分離し、異なる情動状態をもつケース(7)までに至る。カーブ矢印の向きは、情動状態の位相を示し、情動的共感(EE)や認知的共感(CE)(4)までは、同位相で同調している。EEとCEまでに必要な基本となる構造は図2-5-8の左下に示すように調和すべき他者を含む環境との同調である。しかしながら、これ以降は、情動制御応力により、非同調(異なる情動)状態になりうる。

■ 図2-5-8 共感発達モデルとしての自他認知過程



出典:文献 [21] の Fig.6 を改変

同情や哀れみは、このように自他で異なる情動状態を持つ(5)。直感的には、同情は、おもにEE寄りで、哀れみは、主にCE寄りと考えられる。なぜなら、同情はより情動的であり、哀れみは、他者の情動状態を論理的に理解したあとに実感されると考えられるからである。しかしながら、ともに他者の情動状態を知覚し、その要因を理解する必要があり、実際にその差は少ないと察せられる。同調の基本構造に加え、知覚された情動に対する同調への抑制が、エージェンシーの確立(自他識別)にもとづき、必要である(図2-5-8下段中央参照)。

上記の自他における情動状態の不一致(非同調)は、内的と外的の2つの異なる方向に進展する。内的進展は以下である。自己の情動空間が2つに分離し(6)、一つは主観的情動(上)であり、もう一つは客観的情動(仮想化された自己の情動:下)であり、これは、一般的な他者の情動状態の投影と考えられる。客観的自己の情動状態の知覚(知覚された情動)は、客観的な判断なのでCE主体に映る。それに対し、それを感じる自己は、より主観的である(感じられた情動)。外的進展は、自他がそれぞれ集団化することで、各集団内では、すべてのメンバーが同期するが、集団間では異なる。進化の淘汰圧により、集団間が競争的になると、敵対心が生まれる。集団は拡張された自己(もしくは他)である。両方のケースとも、仮想化された自己の創造能力(6)や、図2-5-8の下段右に示す機構にもとづき、自己情動のより強力な調整(7)が、様々な情動を創発するために必要である。

## (2) 身体表現の獲得

自分の身体や運動をどのように表現したり、認識するのかといった身体表現の問題は、自己という概念や自分の運動の所有感覚とも関連する重要な課題である。自分の身体の位置や姿勢などの感覚は、内受容感覚や固有感覚とも呼ばれ、自らの手足や他の身体部分の位置を感知する能力として定義されている。ボディスキーマと呼ばれる身体表現は、生物学的及び人工的なエージェントが固有感覚にもとづいて行動を実行することを可能にする。ロボットなどの人工エージェントによって使用される固有感覚情報は、主に姿勢(及びその変化)に関連し、従来のロボティクスでは、リンク構造の関

節角(及び関節角速度)からなる。これに生物学的に対応する表象は、ボディスキーマやボディイメージと呼ばれ、その区別や定義は定かではなく、論争的である。これらのシステムの基礎をなす神経構造は、現在のイメージング技術の進展により、解明が進んでいるものの、完全ではない。そこで、認知発達ロボティクスアプローチにより、新たな洞察と理解を求める研究が進んでいる。それを紹介しながら、ミラーニューロンシステムの話へとつなげる。

### ① 身体表現の生物学的原理

固有感覚の本来の意味は「自分自身」、または「個々の」という意味のラテン語 *propius* に由来し、また、*capoi*は把握を意味する。したがって、神経科学のみでなく、心理学、また身体の無意識の認識や意識の間の区別と関連して、時には哲学の課題にもなりうる。また、「固有受容」を明確に定義することは困難であり、身体表現、特にボディスキーマやボディイメージの区別も同様である。Head and Holmes [24] は、マルチモーダルな感覚データが統一されている無意識の神経マップとしてボディスキーマを定義し、身体とその機能の明示的な精神的表現としてボディイメージを定義している。前者はしばしば行動と見なされ、後者は知覚のためと見なされている。生物学的システムにおける身体表現は柔軟であり、異なる感覚様式からの情報の時空間統合によって獲得されるという一般的な合意があるが、その構造及び機構の詳細は明らかではない。

可塑性は、身体表現の最も重要な特性の一つである。その起源は、自己身体に触れる反復運動がしばしば観察される子宮内の胎児発達の時期から来る可能性があり、胎児はその運動と結果的感覚との間の関係を知ると考えられている [25]。この初期の表現は、身体の所有感覚や主体感覚(エージェンシー)などの重要な概念とリンクしており、発達後期にボディスキーマやボディイメージに分かれる前の混合体と考えられる。

身体表現における柔軟性及び適応性は、神経可塑性によって引き起こされる望ましい特徴であり、道具使用の場合に観察することができる。Maravita and Iriki [26] は、熊手で食料を取っていたマカクザルによる道具の使用中にボディスキーマが拡張することを発見した。体性感覚刺激と視覚刺激の両方に反応するバイモーダルニューロンと呼ばれるニューロンの活動を脳内皮質から記録した。彼らは、手の体性感覚刺激と手の近くの視覚刺激に応答する「遠位型」と、視覚受容野が手を中心にしていないが手の届く範囲全体に及ぶ「近位型」の2種類のニューロンを見いだした。両方のタイプのニューロンは、道具の使用の経験にも、ツールを使用するサル自身のモチベーションにも適応する。

神経心理学的異常の研究は、ボディスキーマの構築の基礎となるメカニズムと、これらが損傷によってどのように影響されるかを理解するのに役立つ。最も興味深い例の一つは、Ramachandran and Blakeslee [27] によって記述された、いわゆる「幻肢」現象である。四肢欠損のために幻肢の痛みを苦しむ患者は、鏡の箱の無傷の反対側の肢の観察を通じた視覚フィードバックによって痛みを和らげることができ、この経験を通して体の皮質的表現が再構成された可能性があることを示した。ボディイメージとボディスキーマ、障害、及び身体所有とエージェンシー、フォワードモデルなど、他の重要な概念の定義と役割に関するより多くの議論は文献 [28] を参照されたい。

### ② 身体表現の認知発達ロボティクスアプローチ

ロボットのような人工エージェントの場合、一連の動作を実行することによって所定のタスクを達成するためには、身体表現が不可欠である。典型的な(そして伝統的な)状況は、テーブル上のいくつかの物体に到達して把持するために必要とされる関節の数(典型的には3~6)に対応する多数の関

節を電気モーターによって駆動するロボットアームである。対象物のピックアンドプレースの場所とそのサイズ(重量)が固定されて与えられている場合、事前に計画された経路に従ってロボットの肩、上腕、前腕、グリッパーを動かす作業が必要だった。ほとんどの場合、解析的解が与えられ、経路計画は難しくない。ボディスキーマは、各リンクの長さ及び可動範囲情報を用いて、ロボットアームの構成要素がどのように接続されているかを示すリンク構造に対応する。これらのパラメーターは事前に与えられ、関節角が測定され(通常、エンコーダーが使用される)、固有感覚の関数として追従する経路をモニターする。このアプローチは明示的モデル型と呼ばれる。一方で、認知発達ロボティクスアプローチでは、環境との相互作用を通じてこれらのパラメーターを非明示的に推定する。

生物学的な規範は従来のロボティクスの事例とは、以下の点で異なる。

- ロボットのリンク構造は一般的に固定であるのに対し、生物学的なものは柔軟性があり、環境と自己身体の両方の変化に適応する。
- 従来のロボティクスでは、パラメーターを推定するための知識が外部から与えられるのに対し、認知発達ロボティクスアプローチでは、非明示的に推定する。
- マルチモーダルな知覚情報の統合は、生物学的には当然だが、従来ロボティクスは一般的にクロスモーダルの関連づけが含まれていない。

認知発達ロボティクスでは、コンピューターシミュレーションや実際のロボット実験を用いて、神経科学や発達心理学などの分野からの知見をモデル化し、それにもとづいて仮説を検証する。認知発達ロボティクスのアプローチで研究されている身体表現に関連する神経科学的知見の例がある。それは、腹側頭頂間溝領野(VIP、LIP)などの脳領域と神経相関や多様な感覚の連関に関係する時空間偶発性(期待)、不変性、目/頭/頸部間の座標変換などのトピックである。

#### (a) 自己身体の発見

身体表現の課題を扱う前提として、表現の対象となる物理的部位や部分を同定していることが前提となるが、まずは、どのようにして、自己の身体を発見するのであろうか？ 強化学習のマルチエージェントへの拡張[29]では、局所予測モデルを用いて、自分の運動指令によるものも含めて環境の変動を予測するモデルを構築し、状態空間を構成している。その際、センサー空間における自己身体部分の発見アイデアは、「自身が生成した運動指令と相関を持つセンサーデータ部分は、静止環境か自身の身体部に対応する(例えば、手を動かしたときの手の視覚映像、頭を右に振れば、画面全体が左に流れるなど)」であった。静止環境との区別は、重力方向などの事前知識などにより可能とされていた。

#### (b) 道具使用による適応的身体表現

Hikitaら[30]は、マカクザルの道具使用による適応的な身体表現の計算モデル化を試みた。視覚、触覚、及び固有感覚からクロスモーダル身体表現を構成する方法により、適応的な身体表現を可能にした。ロボットが何かに触れると、触覚の活性化は、顕著性マップにもとづいて視覚的に見いだされ、結果としてエンドエフェクターとみなされる身体部分の視覚受容野の構築過程を引き起こす。同時に、固有感覚情報が、この視覚受容野に関連づけられて、クロスモーダルの身体表現を構成する。コンピューターシミュレーションと実際のロボット実験の結果は、マカクザルに見られる頭頂のニューロンの活動[31]に対応する活動を示した。

### (c) VIPニューロンの働き: 頭部身体周辺空間の表現の獲得

Maravita and Iriki [26] は、道具使用時に、頭頂葉の腹側頭頂間溝領野 (VIP、LIP) に視覚と触覚に反応するバイモーダルニューロンの活動の知見を紹介したが、これは、生物の身体表現が動的に構成されている可能性を示している。よって、人間は、随時経験から視空間内に存在する物体の位置を表現するための様々な参照枠 (身体中心参照枠、物体中心参照枠) の概念を獲得し、さらにはそれらにもとづいて表現された位置と触覚や体性感覚などの異種感覚を柔軟に統合することで、身体表現を獲得していると考えられる。Fukeら [31] は、エージェントが自身の視触覚経験を通して頭中心参照枠での視空間表現だけでなく、自身では直接観測不可能な顔部位の視触覚表現を学習するモデルを提案している。

### ③ 身体表現の課題

身体表現は、現実世界で働く自然システムと人工システムの最も基本的な問題の一つであるが、これまで、それぞれ個別に考えられてきた。近年、義肢義足などのハイブリッドシステムとして動作することもあり、今後、より親密に関連し合うだろう。本項では、生物学とロボット学の両方における身体表現の研究を簡単に概説した。しかし、身体表現がどのようなメカニズムの下でどのように作用するかは、完全には明らかではなく、さらなる研究推進が望まれる。

生態学的自己の次は対人的自己の形成であるが、ここが社会的相互作用を通じた認知の発達過程である。図2-5-9に示すように、対人的自己の形成には、ミラーニューロンシステム (以降、MNSと略記) が重要な役割を果たす。このMNSの発達過程も含めて、社会的相互作用による認知発達のモデル化を試みる予測符号化によるアプローチを次節では紹介する。

■ 図2-5-9 予測学習にもとづくロボットの認知発達モデル



出典: 長井志江「認知発達の原理を探る: 感覚・運動情報の予測学習に基づく計算論的モデル」[32]を改変

### (3) 予測誤差最小化原理による社会的相互作用創発

社会的認知機能の発達を統一的に説明する理論として、感覚・運動情報の予測学習にもとづく計算モデルが導入されている [33]。認知発達の様々な側面を統一的に扱おうとする試みである (図2-5-10)。

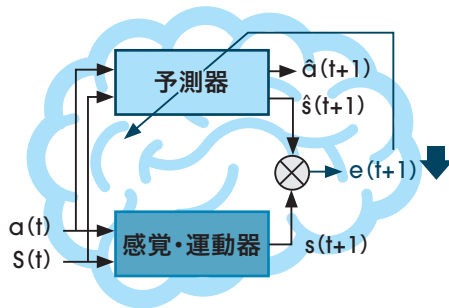
予測学習とは、身体や環境からのボトムアップな感覚信号と、脳が内部モデルをもとにトップダウン

ンに予測する感覚信号の誤差を最小化するように内部モデルを更新したり、環境に働きかけたりすることである。認知発達過程に適用する際、2つの過程が考えられる。一つは、自己の感覚・運動経験を通じた予測器の更新による予測誤差の最小化過程で、自他認知、目標指向動作などが含まれ、認知発達初期過程に対応する(図2-5-10(a))。それに対し、社会的な環境では、予測した運動の実行による他者運動起因の予測誤差の最小化が課題で、模倣や援助行動などが対象である(図2-5-10(b))。以下では、前者の例として、ミラーニューロンシステムの発達を、後者の例として、他者行動の予測から利他的行動に至る過程を紹介する。

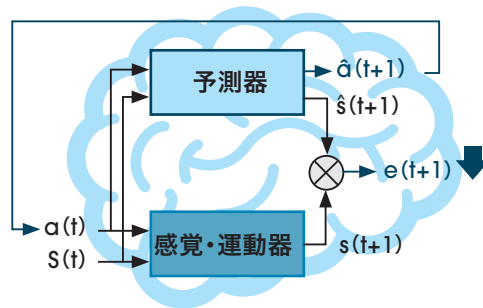
■図2-5-10 予測誤差学習



(a) 自己の感覚・運動経験を通じた  
予測器の更新による予測誤差の最小化  
→ 自他認知、目標指向動作など



(b) 予測した運動の実行による  
他者運動起因の予測誤差の最小化  
→ 模倣、援助行動など



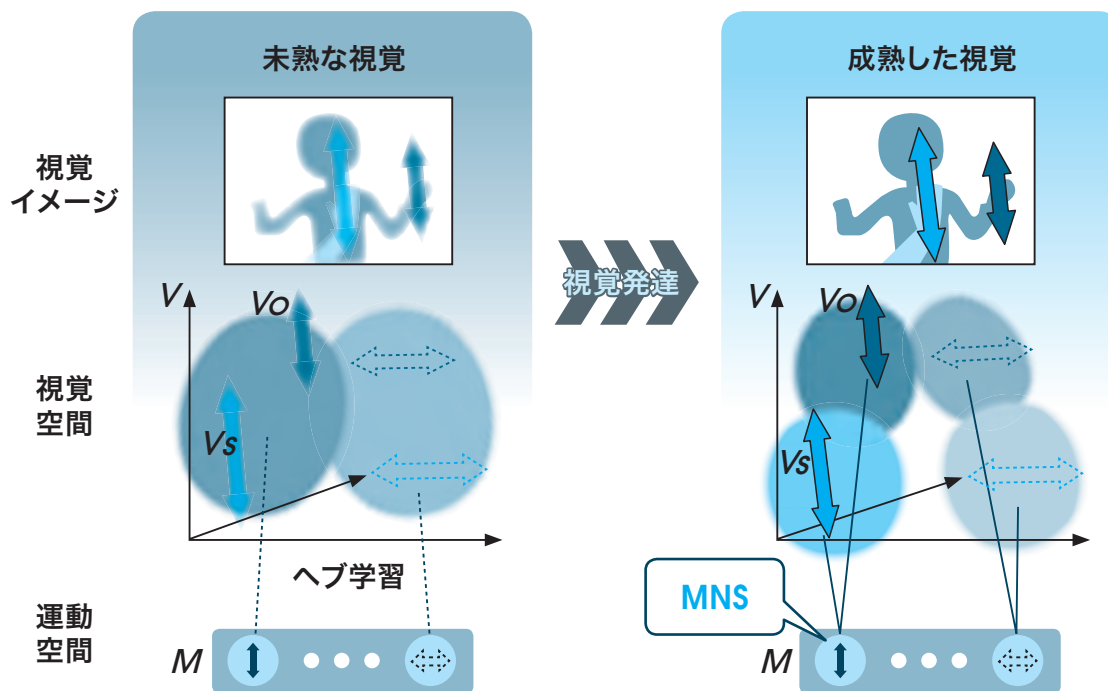
### ① MNSの発達

MNSについては、『AI白書2017』の「1.3.3.2 ミラーニューロンシステムと社会性発達基盤」に詳しい。簡単に言えば、自身の行為の実行と他者による同一行為の観測を同一のニューロンが符号化しており、これがマカクザルのF5と呼ばれる運動前野で発見された[33]ことに由来する。人間も同様のシステムを有しているのではないかと、システムと称されている。

MNSが生得的か生後の環境での学習・発達によって獲得されるかの議論は尽きない。認知発達ロボティクスの立場としては可能な限り後者の立場をとり、説明可能なモデルを構築してきた。MNSについては、Kawaiら[34]が、発達モデルを提唱している。視覚発達を伴う感覚運動学習がMNS発達を促進するという仮説の下に、未熟な感覚の時期には自他が混同されるが、発達に伴い自他分離が起り、これがMNSの基盤となっているという主張である。言語の習得をはじめとして、乳幼児が未熟であるがゆえに、様々な認知機能の学習を促進していることが示唆されているが[35]、その計算モデル化の一つと言えよう。もう一つ、養育者側がそれを明示的・非明示的にサポートしていることも忘れてはならない。MNS創発の場合、他者(養育者)が、赤ちゃん(ロボット)の真似をしてくれることが前提となっている。

図2-5-11に基本アイデアを示す。他者運動の観測で得られるオプティカルフローベクトル(両矢印)は、ガウス関数を用いて空間的にフィルタリングされる。フィルターの幅は視空間Vにおける自己組織化の進展によって決定され、徐々に狭まる。発達の初期段階(左)では、ロボットはモーションコマンドとオプティカルフローとを関連づけるが、画像が粗いため、自身のフローと他者のフローを区別できない。発達後期(右)では、狭いフィルターにより、自己と他者のフローの区別が可能になり、これがMNSの基盤をなす。

■図2-5-11 視覚発達による MNS 創発モデル:(左) 発達初期, (右) 発達後期



出典: Yukie Nagai「Emergence of Mirror System Immature vision leads to self-other correspondence」のFig.3を改変

## ② 他者視点取得の困難さ

MNSが発達したことで、相手の行動観察から、自己の運動が励起され、そのことにより相手の行動の目的と意図が理解可能になると期待される。上記の例では、簡単な運動に限っていたので、識別が容易であったが、複雑な動きに対しては、どのように考えられるだろうか。自他分離が可能になったことで、図2-5-8にある「自己の気づき」の段階に入ってきた。そして次の段階として「他者視点取得 (Perspective Taking)」の課題が上がってくる。すなわち他者の視点に立てるかという課題である。同一行動の主体の差異(自己運動か他者運動)による観察の見かけの違いの吸収で、座標変換の課題とも捉えられる。頭頂葉で自己座標系と他者座標系の変換が行われているようだ[36]が、発達の観点から、生得的と考えるよりも生後の学習の結果として変換プログラムが構築されたと見なしたい。とすれば、いかにして可能か。サルの場合のゴール指向の他動詞的動作であれば、強化学習のスキームで報酬獲得による価値の等価性[37]により、同一行動の異なる視点からの観察による見かけの違いばかりでなく、実現方法が異なる行動でも等価とみなすことで、結果として、座標変換が可能と考えられる。ただし、これはサルの場合に対応し、ヒトの場合は、より一般的なスキームが必要かもしれない。例えば、物体操作など学習や発達を通じて、半ばゴール指向、半ば視触覚融合の連

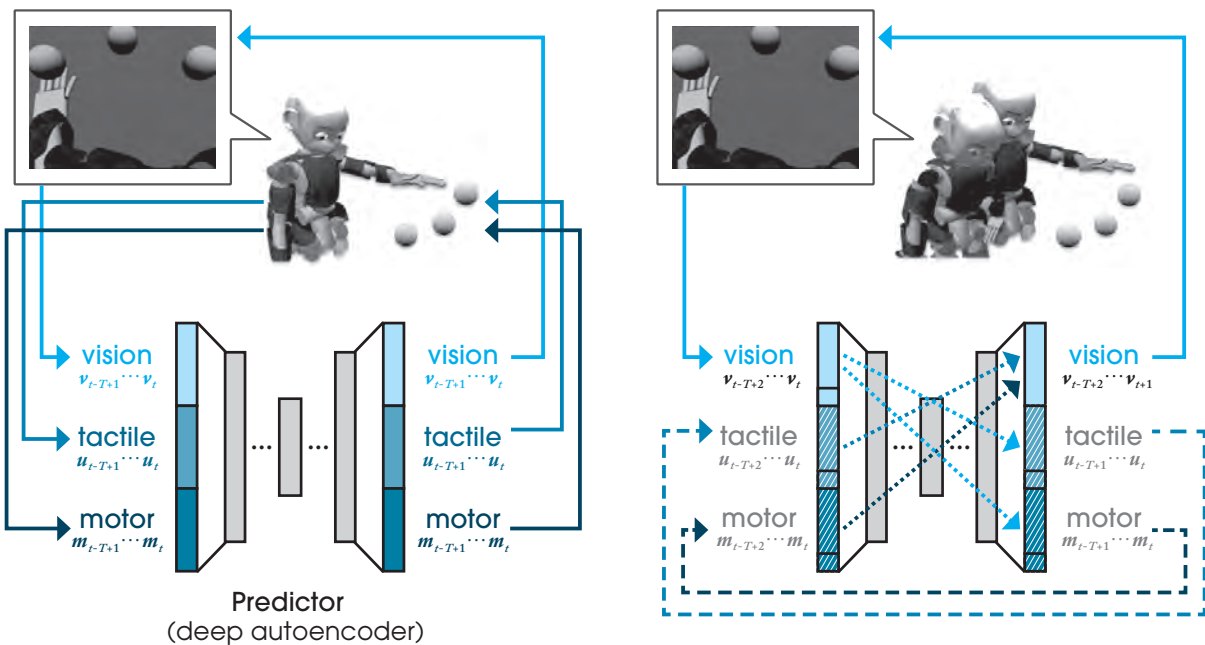
統的表象構築により、結果として座標変換が可能かもしれない。その際には、並行する他の認知機能の発達との兼ね合い(独立か、相互促進か、干渉か)も、構成的手法の観点から興味深い課題である。

### ③ 他者運動に起因する予測誤差の最小化による援助行動やターンテイキングの創発

まず、自己の運動生成経験にもとづく他者運動の予測課題では、最初に自己の運動生成経験を通じた感覚・運動信号の予測学習を実施し、その後、自己の運動経験にもとづく他者運動の予測が可能になる。オートエンコーダーを用いた手法を図2-5-12に示す[38]。左が前者で右が后者である。後者の過程では、先に示した他者視点取得を前提としており、今後、この課題をどのように解決するかが課題である。

次に、社会的行動の一種である利他的行動の創発について考えよう。発達心理学の分野では、14カ月の乳幼児が利他的行動を示すといわれている[39]。そこでは、乳幼児は社会的信号(視線や発話)や、利他的行動に対しての報酬を受けずに、自発的に実験者を助けようとする。この過程が先に示した自己の予測器を利用した他者運動の予測と、他者起因の予測誤差を引き金とした運動の生成に対応し、これが結果として利他的行動に映ると考えられ、その過程の計算モデル化が行われた[40]。

■ 図2-5-12 深層型オートエンコーダーを用いた複数感覚・運動信号の予測学習

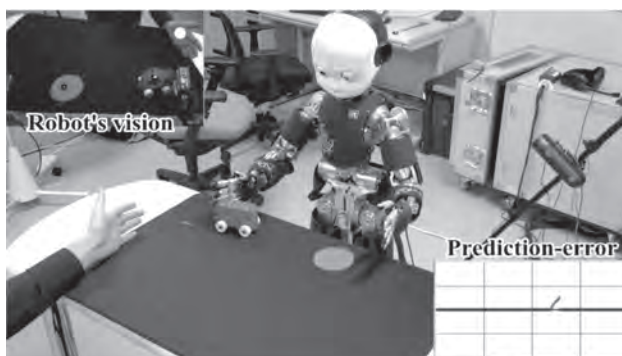


出典:文献[38]をもとにJorge L.Copete,Dr.Yukie Nagaiが作成

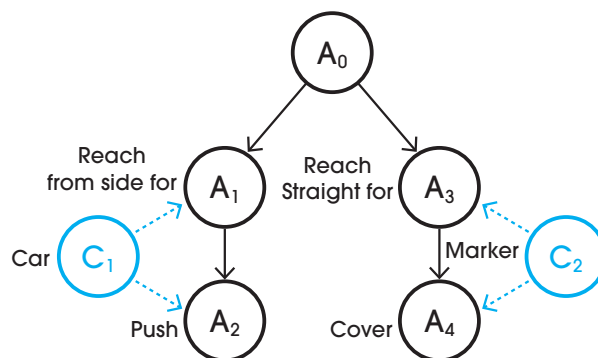
図2-5-13(a)に実験の様子を、(b)に感覚・運動の確率遷移モデルで表現したロボット内部モデルを示す。最初に青い車に対しては左側にプッシュ、赤い円盤に関しては、それを手で覆って隠す行為をそれぞれの物体に対するアフォーダンスとして学習する。学習後に実験者が青い車を押そうとするが押せずにいると、アフォーダンスから生じる誤差が増大し(図2-5-13(a)の右下)、その誤差を解消しようとして、青い車を押す行為が生じる。その結果、ロボット自身が意図せず(意図したものは誤差の解消のみ)、他者への利他的行動を創発したことになる。



■ 図2-5-13 利他的行動の創発



(a) ロボットによる利他的行動の創発実験



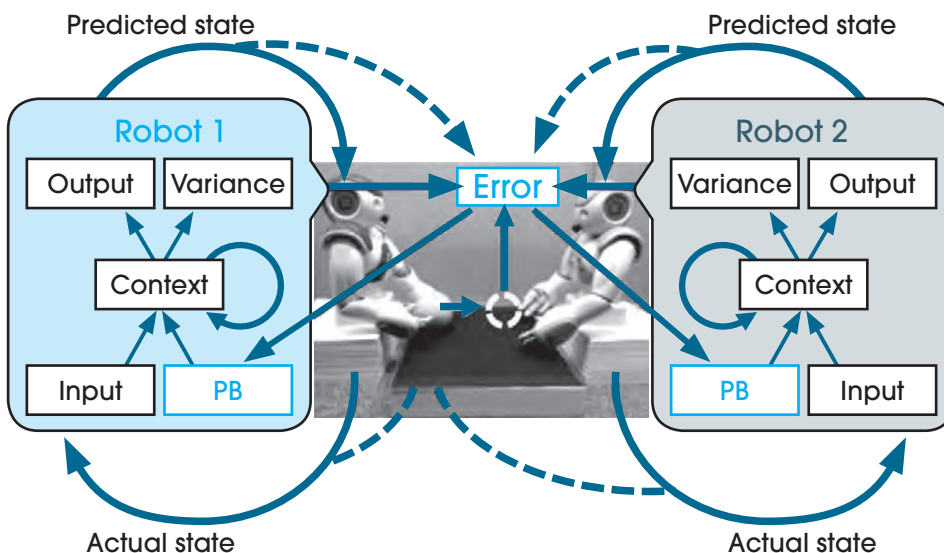
(b) 感覚・運動の確率遷移モデルとして表現されたロボットの予測器

出典: 文献 [40]より引用

尾形らは、予測誤差最小化原理によるコミュニケーション創発実験を行っている [41]。運動学習を終えた2体のヒューマノイド間でボールのやり取りを通じて、時系列を覚える再帰型NNによる予測誤差最小化が、ターンテイキングなどの社会的相互作用を自律的に生じさせた (図2-5-14)。

上記の例は、予測誤差最小化原理の一般性を強く示しており、構成的発達科学の学習規範として有望であり、後者では二者間でのプロセス共有がコミュニケーションの大きな要素であるが、そのことや自己・他者などへの気付きが、今後課題となろう。図2-5-8に示したこれらの心的機能の総体が意識として捉えられるのではないかと考えられる。次項では、本節の締めくくりとして、人工意識の設計に関して神経科学的側面からの見方を紹介し、最後に今後の展望を述べる。

■ 図2-5-14 実験構成: 実線は体性感覚、破線は視覚



出典: 文献 [41]より引用

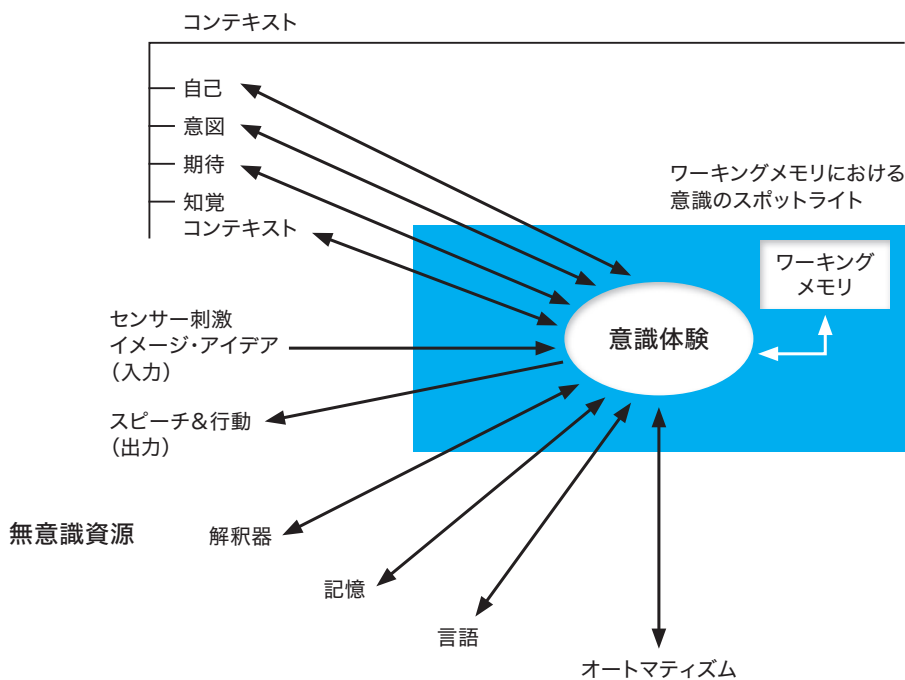
## 2.5.5 人工意識に関する認知神経科学的考察

具体的に人工意識を実現しようとするとき、まずは、人間の意識の機構的な構造の解明は設計指針のみならず、検証の意味でも重要である。ここでは、認知神経科学者のDehaeneの著書[42]や彼らの論文[43]を交え、人工意識をどのように捉えるかを議論する。

Dehaeneは、認知神経科学的に、意識と無意識を定義し、それらの関係から、主観的な現象を尊重することで、意識的知覚を実験によって操作できるということを発見し、様々な実験を通じて意識の解明を目指してきた。無意識を分類し、意識に最も近い前意識が数多く並列で作動しており、注意のボトルネックを通じて、ポップアップされる機会(バイズ推定)を窺っている。一旦ポップアップされて意識レベルに昇ると、系列的処理の集中を受け、その他の無意識の知覚がマスクされる。この構造はBaars[44]が提案した大局的作業空間理論(Global workspace theory、以降GW理論と略記)と合致し、グローバル・ニューロナル・ネットワークとして提案されている。GW理論は、AI分野で著名なNewellらのブラックボードアーキテクチャーに基盤がある。分散した一連の知識ソースから構成され、単一の構成要素だけでは解決できない問題を共同で解決することを示した。GW理論は、知覚、感情、動機づけ、学習、作業記憶、自発的制御、及び脳における自己システムの意識的側面について明示的な予測を生成する。これは、神経ダーウィニズムや脳機能の力学理論などの生物学的理論と類似している。図2-5-15にその概念図を示す。意識内容は、ワーキングメモリ上のスポットライトに対応し、無意識の脳領域(意識経験を直接サポートしないと考えられる皮質領域、海馬、及び基底核)と関連する。意識的認知そのものは、常に無意識の文脈によって形成され、エグゼクティブ・ファンクション(自己)は、そのようなコンテキストの一つのセットと考えられる。

Dehaeneの定義に従えば、原理的には乳幼児や霊長類、さらには、他の哺乳類や動物にも意識は生じる。ただし、ヒトの場合、言語がコミュニケーションシステムとしてよりも、表象装置として進化し、考え出す能力を与えたとして、表象レベルが他の動物に比べ抜き出ているという。

■ 図2-5-15 GW理論の概念図



出典:文献[44]のFig.1を改変

Dehaeneら [43] は、機械の意識を検討するうえで、明確に意識を所有していると思われるヒトの脳について再考し、機械が創発する意識のありさまを議論している。脳の中には、2つの情報処理計算が存在すると主張する。それらは、以下のとおりである(図2-5-16)。

1. 大局的情報伝達のための情報選択 → それを計算と報告に柔軟に利用可能にすること(第一センスの意味でC1意識と呼ぶ)。
2. 上記の計算の自己モニタリング → 確信かエラーかの主観的感覚への導入(第二センスの意味でC2意識と呼ぶ)。

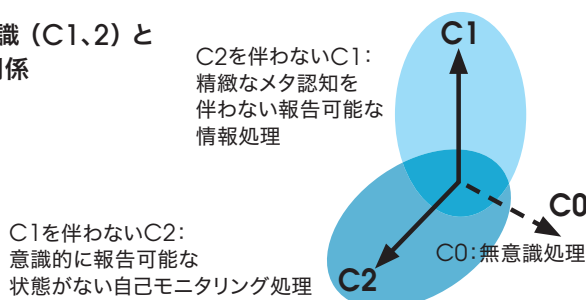
現状の機械は、脳の中でいうところの無意識処理(C0意識と呼称)に対応しているとみなす。C0は、心ない自動的な応答により、精練された情報処理を行っている。顔や音声の認識、チェスのゲーム評価、構文解析などを無意識に、すなわち大域情報伝達や自己モニタリングなしに行っている。それらは、独立したモジュールであり、個々の深層学習ネットワークに対応する。

C1は大局的に情報収集し、単一モジュールでは解決できなかった課題に対して、柔軟に対応可能であり、先に挙げたGW理論が対応する。C1が外界の情報収集に専念していたことに対し、C2では、自己の内部を反映する情報処理構造が特徴であり、それゆえ、彼らは、C1とC2が直交していると考えている。当然、我々の日常生活では、C0、C1、C2すべてが混在するのであるが、C2を伴わないC1やC1を伴わないC2が存在することで、その直交性を担保している。

現状の多くのシステムがC0状態であるとする、機械の意識を実現するうえでは、C1やC2レベルの実現を検討しなければならない。C1レベルはすでに紹介したGW理論に則ったものが構築中であるが、C2レベルではどうであろうか。Dehaeneらは、一つの有力なメカニズムとして、敵対的生成ネットワーク(Generative Adversarial Networks、略称: GANs)を挙げた。これは、生成ネットワーク(generator)と識別ネットワーク(discriminator)の2つのネットワークから構成され、前者の出力を後者がその正否を判定する。生成側は識別側を欺こうと学習し、識別側はより正確に識別しようと学習する。これは、自己モニタリングの機能に相当すると考えられたからである。結論として、意識は情報計算の産物であり、C1とC2の能力を持つ機械であれば、何かを見、それに対する自信を表明し、他者に報告できることを知っている。そのため、モニタリング機構が故障すると幻覚状態に陥り、人間と同じような錯覚を経験するとしている。

これまで見てきたように、認知神経科学の側面からの意識の脳内現象の構造的理解は進んできたと思われるが、C2レベルのより深い構造、すなわち自己の内省のみならず、2.5.3(1)で示した主観と客観が混在し、自他が混じり合った状況で垣間見られる心のダイナミクスの機構には触れられていない。それに対し、2.5.4などで議論されている身体性やその身体に痛み回路が想定されることで、共感や倫理観の創発の可能性も議論可能である[19]。

■ 図2-5-16 直交する二種の意識(C1,2)と無意識(C0)の関係



出典: 文献 [20] の図3

## 2.5.6 おわりに

本節では、「身体性とロボティクス」と題し、ロボティクスの学術分野としての意味合いをAIのコンテキストから論じた。深層学習を核とする現在の人工知能の分野の流れを考えると以下の2つが浮き上がる。

- 1.現在の深層学習を徹底的に突き詰め、ビッグデータをベースにend-to-endの学習機構により、意味をグランドすることなく、使えるシステムをどんどん社会に出していき、有用なツールとして磨きをかける。
- 2.本節で論じた身体性、最後に言及した痛みの神経回路をロボットに賦与することで、MNSを通じた共感行動生成、さらには、倫理観の創出など、人間自身の認知発達過程のミステリーを構成的に解きほぐすことで、よりグランドした人工システムの設計から実証実験を繰り返す。

前者がやや工学的、後者がやや科学的アプローチと言えなくもないが、現実や将来は、互いに影響を及ぼし合いながら、突き進んでいくものと察せられる。未来共生社会を描くとき、これらが新たな社会創造の価値観につながると期待する。

### ◆参考文献

- [ 1 ] 日本ロボット学会(編)『新版ロボット工学ハンドブック』コロナ社, 2005.
- [ 2 ] Ian Lenz, Honglak Lee, and Ashutosh Saxena. Deep learning for detecting robotic grasps. CoRR, Vol. abs/1301.3592, 2013.
- [ 3 ] Joseph Redmon and Anelia Angelova. Real-time grasp detection using convolutional neural networks. CoRR, Vol. abs/1412.3128, 2014.
- [ 4 ] Lerrel Pinto and Abhinav Gupta. Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours. CoRR, Vol. abs/1509.06825, 2015.
- [ 5 ] Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. CoRR, Vol. abs/1504.00702, 2015.
- [ 6 ] 細田、浅田。「構造やパラメーターに関する先験的な知識を必要としないフィードフォワード補償器を持つ適応型ビジュアルサーボ系の構成」。日本ロボット学会誌, Vol. 14, No. 2, pp. 159-165, 1996.
- [ 7 ] Frederik Ebert, Chelsea Finn, Alex X. Lee, and Sergey Levine. Self-supervised visual planning with temporal skip connections. Proc. of 1st Conference on Robot Learning (CoRL 2017), 2017.
- [ 8 ] Jun Tani. Exploring Robotic Minds: Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena. Oxford University Press, 2016.
- [ 9 ] Chelsea Finn, Ian J. Goodfellow, and Sergey Levine. Unsupervised learning for physical interaction through video prediction. In Neural Information Processing Systems (NIPS), 2016.
- [ 10 ] Chelsea Finn and Sergey Levine. Deep visual foresight for planning robot motion. Proc. of 2017 International Conference on Robotics and Automation (ICRA), 2017.
- [ 11 ] 稲谷龍彦. 技術の道德性と刑事法規制. 松尾陽(編)『アーキテクチャと法』第4章, pp. 93-128. 弘文堂, 2017.
- [ 12 ] エトムント フッサール (著), 浜渦辰二 (翻訳)『デカルト的省察』岩波文庫, 2001.
- [ 13 ] マルティン ハイデッガー (著), Martin Heidegger (原著), 細谷 貞雄 (翻訳)『存在と時間〈上〉』ちくま学芸文庫, 1994.
- [ 14 ] マルティン ハイデッガー (著), Martin Heidegger (原著), 細谷 貞雄 (翻訳)『存在と時間〈下〉』ちくま学芸文庫, 1994.
- [ 15 ] モーリス メルロ=ポンティ (著), 竹内 芳郎 (翻訳), 小木 貞孝 (翻訳)『知覚の現象学 1』みすず書房, 1967.
- [ 16 ] モーリス メルロ=ポンティ (著), 竹内 芳郎 (翻訳), 木田 元 (翻訳), 宮本 忠雄 (翻訳)『知覚の現象学 2』みすず書房, 1974.
- [ 17 ] ブルーノ ラトゥール (著), Bruno Latour (原著), 川崎 勝 (翻訳), 平川 秀幸 (翻訳)『科学論の実在一バンドラの希望』産業図書, 2007.
- [ 18 ] ピーター=ポール フェルベーク (著), Peter-Paul Verbeek (原著), 鈴木 俊洋 (翻訳)『技術の道德化: 事物の道德性を理解し設計する』法政大学出版局, 2015.
- [ 19 ] 浅田稔「痛みを感じるロボットの意識・倫理と法制度」人工知能学会誌, Vol. 33, No. 4, pp. 450-459, Jul 2018.

- [20] M. Merleau-Ponty. *The Visible and the Invisible: Followed by Working Notes* (Studies in phenomenology and existential philosophy). Northwestern University Press., 1968.
- [21] Minoru Asada. Towards artificial empathy. *International Journal of Social Robotics*, Vol. 7, pp. 19–33, 2015.
- [22] Ulric Neisser, editor. *The Perceived Self: Ecological and Interpersonal Sources of Self Knowledge*. Cambridge University Press, 1993.
- [23] Minoru Asada. Can cognitive developmental robotics cause a paradigm shift? In Jeffrey L. Krichmar and Hiroaki Wagatsuma, editors, *Neuromorphic and Brain-Based Robots*, pp. 251–273. Cambridge University Press, 2011.
- [24] H. Head and H. G. Holmes. Sensory disturbances from cerebral lesions. *Brain*, Vol. 34, No. 2–3, pp. 102–254, 1911.
- [25] Philippe Rochat. Self-perception and action in infancy. *Experimental Brain Research*, pp. 102–109, 1998.
- [26] A. Maravita and A. Iriki. Tools for the body (schema). *Trends Cogn. Sci.*, Vol. 8, No. 2, pp. 79–86, 2004.
- [27] V. S. Ramachandran and Sandra Blakeslee. *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. Harper Perennial, 1998.
- [28] Matej Hoffmann, Hugo Gravato Marques, Alejandro Hernandez Arieta, Hidenobu Sumioka, Max Lungarella, and Rolf Pfeifer. Body schema in robotics: A review. *IEEE Transactions on Autonomous Mental Development*, Vol. 2, No. 4, pp. 304–324, 2010.
- [29] Minoru Asada, Eiji Uchibe, and Koh Hosoda. Cooperative behavior acquisition for mobile robots in dynamically changing real worlds via vision-based reinforcement learning and development. *Artificial Intelligence*, Vol. 110, pp. 275–292, 1999.
- [30] Mai Hikita, Sawa Fuke, Masaki Ogino, Takashi Minato, and Minoru Asada. Visual attention by saliency leads cross-modal body representation. In *The 7th International Conference on Development and Learning (ICDL'08)*, p. to appear, 2008.
- [31] Sawa Fuke, Masaki Ogino, and Minoru Asada. Acquisition of the head-centered peri-personal spatial representation found in vip neuron. *IEEE Transactions on Autonomous Mental Development*, Vol. 1, No. 2, pp. 131–140, 2009.
- [32] 長井志江. 認知発達の原因を探る: 感覚・運動情報の予測学習にもとづく計算論的モデル. *ベビーサイエンス*, Vol. 15, pp. 22–32, Mar 2016.
- [33] Rizzolatti G., Camarda R., Fogassi M., Gentilucci M., Luppino G., and Matelli M. Functional organization of inferior area 6 in the macaque monkey: II. area f5 and the control of distal movements. *Exp. Brain Res.*, Vol. 71, pp. 491–507, 1988.
- [34] Yuji Kawai, Yukie Nagai, and Minoru Asada. Perceptual development triggered by its self-organization in cognitive learning. In *Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5159–5164, 2012.
- [35] E. L. Newport. Maturational constraints on language learning. *Cognitive Science*, pp. 11–28, 1990.
- [36] K. Ogawa and T. Inui. Lateralization of the posterior parietal cortex for internal monitoring of self-versus externally generated movements. *Journal of Cognitive Neuroscience*, Vol. 19, pp. 1827–1835, 2007.
- [37] Yasutake Takahashi, Yoshihiro Tamura, and Minoru Asada. Mutual development of behavior acquisition and recognition based on value system. In *Proceedings of the 10th international conference on simulation of adaptive behavior (SAB08)*, pp. 291–300, 2008.
- [38] Jorge L. Copete, Yukie Nagai, and Minoru Asada. Motor development facilitates the prediction of others' actions through sensorimotor predictive learning. In *Proceedings of the 6th IEEE International Conference on Development and Learning, and Epigenetic Robotics (ICDL-EpiRob 2016)*, pp. (CD-ROM), 2016.
- [39] Felix Warneken, Frances Chen, and Michael Tomasello. Cooperative activities in young children and chimpanzees. *Child Development*, Vol. 77, No. 3, pp. 640–663, 2006.
- [40] Jimmy Baraglia, Yukie Nagai, and Minoru Asada. Emergence of altruistic behavior through the minimization of prediction error. *IEEE Transactions on Cognitive and Developmental Systems*, Vol. 8, No. 3, pp. 141–151, 2016.
- [41] 尾形哲也. 神経回路モデルとコミュニケーションの創発. *日本ロボット学会誌*, Vol. 35, No. 3, pp. 195–198, April 2017.
- [42] スタニスラス・ドゥアンヌ (著), 高橋 洋 (翻訳) 『意識と脳—思考はいかにコード化されるか』 紀伊國屋書店, 2015.
- [43] Stanislas Dehaene, Hakwan Lau, and Sid Kouider. What is consciousness, and could machines have it? *Science*, Vol. 358, pp. 486–492, 2017.
- [44] Bernard J. Baars. Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in Brain Research*, Vol. 150, pp. 45–53, 2005.

## 2.6 ▷ AIと社会

### 2.6.1 社会システムデザインの必要性

インターネットの誕生は社会の在り方を根本的に変えつつある。広範囲の情報を即座に入手できるということが利便性だけではなく、情報処理技術 (IT) の適用範囲を飛躍的に増大させている。しかしながら、ITの潜在能力はフルには発揮されていない。今後はこれにAI技術が加わり能力はさらに向上することが期待されているが、使う側が全く追いついていないのが現状である。

公的機関への申請手続きに関しては、オンライン化、マイナンバーによる添付書類の削減などが進められているものの十分とはいえず、特に個人の負担は大きい。例えばマイナンバーカード。個人認証はオンラインで可能なのに、講演料や謝金が発生した場合に、ほとんどの組織が郵送でマイナンバーの提出を求めている。そして年度末には紙の源泉徴収票が送られてくる。こちらもオンラインで処理ができるはずである。すべての収入にマイナンバーが紐づけられている職業 (ホワイトカラー中心) では確定申告も自動化できるはずである (現状では源泉徴収票を個人が書き写して申請するという原始的なシステムが採用されたままである)。これは国全体としては膨大な資源と時間を無駄にしていることになる。年金についても制度や請求手続きは複雑極まりないし、指定難病患者への医療費助成制度の申請はマイナンバーにより住民票や課税証明書の添付が省略されたものの難病を抱えた本人が行うことは到底困難である。近年のAIによる音声認識や言語生成を活用することで、複雑な制度や手続きや、高齢者や難病・障がいを抱えた個人の申請に対して会話方式で補助・支援することが期待される。

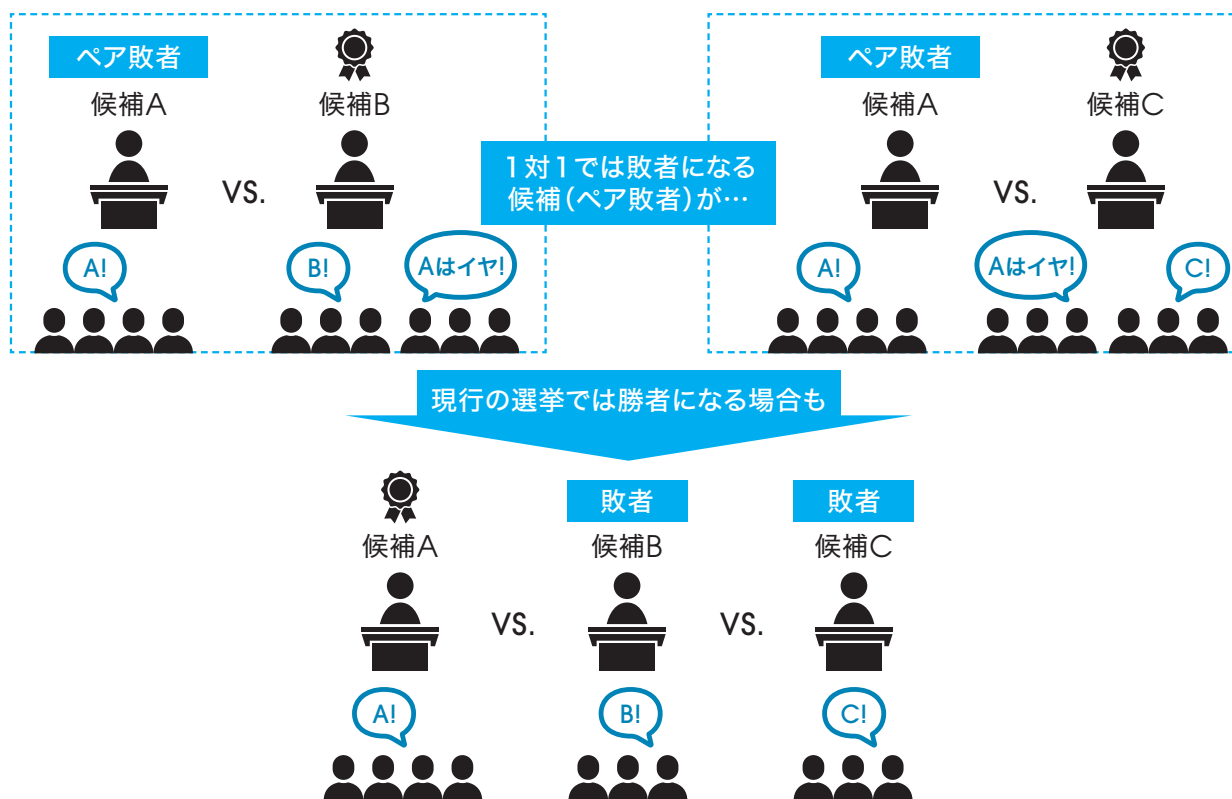
以下の節ではAI技術応用で変革できる様々な社会システムのデザインを例示したい。ただし、これは可能性の列挙であって、実際にそうするのが良いかは専門家 (社会学者等) の研究と判断を必要とする。

### 2.6.2 社会的意思決定システム

民主主義のための仕組み (投票システム、代議員制等) はインターネット以前の形態がそのまま使われている。例えば現在の選挙システムは必ずしも民意を反映しないということは理論的には分かっている [1] にもかかわらず改善されない。投票所に足を運び、投票用紙に記入させるから、投票用紙に政治家が一人しか書けない、小選挙区でも政党一つしか書けない。この方式では、例えば票割れにより民意を反映しない人や政党も選ばれうるため、様々なシステムが研究されている。ITを使えば様々な工夫が可能である。

ベストの投票方式というのは自明ではないが、これだけは避けたいということが自明な方式は分かっている。例えばペア敗者という概念がある。候補者2人を比べ、どちらが良いかと比べた場合に他のすべての候補に負ける候補のことであり、これを選ばないことが重要である。ところがこれが選ばれる可能性のある投票方式がいくつかあって、現在行われている多数決はその一つである (図2-6-1)。

■ 図2-6-1 民意の反映という観点での現行システムの課題



ボルダールールという方式はN人を選ぶ場合に、各投票者が1位にN点、2位にN-1点、N位に1点というように点数をつけ、その合計点で争うものだが、この方式だと少なくともペア敗者が選ばれることはない。逆にペア勝者がいるなら文句なく選びたいところであるが、ペア勝者がいない場合(通常は明らかな勝者はいない)でも通用するような良い方式はなかなかないようである。

政党を選ぶときにマニフェストを判断基準にする。しかし、これは様々な政策が抱き合わせで書いてあるわけなので、個々の政策に対して個別に投票できるわけではない。これは商法で規制されている抱き合わせ販売と何ら変わらない。

AIを使えばもっと良い方法が実現できるはずである。代議士に頼らない方法も考えられる。学問的に新しい、ダイレクトに民意を反映できるようなシステムというのが考えられるべきである。可能性の一つは直接民主制で、インターネットで全国民が議論するというものだ。(現状ではまだ無理であるが)議長はAIが行い、現状の意見や対立点、それぞれの長所と欠点などの明示も可能である。

このように、技術は可能性を広げるが、それらの可能性の中から良いものを選ぶには技術以外の判断基準が必要で、社会学あるいは社会の制度設計をしている人たちが社会デザインに参加すべきである。

### 2.6.3 会社組織と働き方

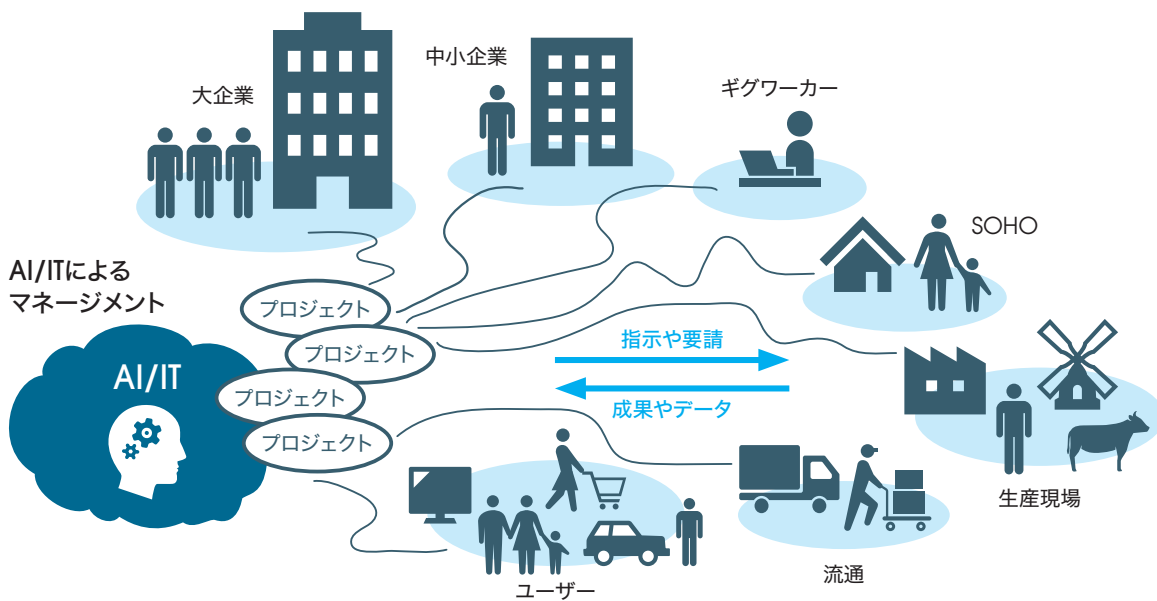
企業ではAIを新しいサービスや製品に取り込むことを検討しているが、企業の仕組みそのものに適用しようとする例はまだ少ない。一部採用面接などにAIを使い始めた企業はあるが、人事全般(メンタルケアなどの部分は除く)、経理、あるいは戦略策定の補助などに使い始めたという話は聞かれない。

組織のマネージメントはAI利用で大幅に変えられるし、変えねばならない。カーツワイルが言う

ように技術の進歩は加速している。これからは破壊的イノベーションがいつ起こるか分からない時代になる[2]。そのような大きな変化に対応するためには現在の、固定の社員が、部とか課という固定の組織に属して仕事をしているような固定の組織では追いつかない。組織やメンバーを必要に応じて組み替えることが必須となろう[3]。社員は自分の所属する部署で仕事をするというのが今の企業形態であるが、AIを使えばその必要がなくなる。ミッション(仕事)ができたときに、それにかかわる人材をその場で、インターネットで集めて管理するということが可能になる。報酬の体系、命令伝達系統など様々な問題はコンピューター(AI)で管理したほうが良い。ある人の仕事やグループを変更することによる影響は大局的なので人間が管理するには限界がある。AI/ITで管理することによって柔軟なマネジメントができる。複雑なシステムの整合性を維持するというのはITの得意分野の一つである(図2-6-2)。

働く側から見た場合にも自由度が増える。会社組織に属することなく様々なプロジェクトに参加し、自分の技能を活かすのだ[4]。現在働き方改革として、育児休暇とかワークシェアとかが議論されている。柔軟な組織体制を導入することによって、フルに働いている状態と育児休暇の間の制度的区別というのはなくなって、単に1日のうち、何時間どこで働くかという問題にすぎなくなる。

■図2-6-2 AI/ITによるマネジメントが変える働き方の変化



いずれにしても今の形態の企業というのは、近いうちになくなるであろう。カリフォルニアではすでに始まっているが、日本の大企業のほとんどは組織体制を変えようとしていない。時代の変化に追従できない組織は数年で(早ければ2020年ごろ、遅くとも2030年ごろまでに)なくなるであろう。

## 2.6.4 経済システム

資本主義は、歴史的に見ると富の集中を促しており、その傾向は近年ますます顕著になっているという趣旨の本が世界中でベストセラーになった[5]。富の再配分をするシステムを作らなければならない。

最近、AIに仕事を奪われるという議論がある。しかし、仕事と収入が1対1に対応しているからこういう議論になるのであって、奪われて困るのは収入のほうだ。仕事をしないで収入を得られるの



なら、あるいは、仕事をしないで暮らせるなら、そのほうが良い人も少なくないであろう。AIを導入すれば全体の生産効率は上がる。経営者は効率が上がる場合しか使わない。そうすれば基本的には我々が働かなければならない総時間は減るはずである。一部の人が総取りをするという今の資本主義のままだと、生活が立ち行かない人が多出して困る。再配分システムのデザインが必要で、ベーシックインカムというのもそういう制度の一つである[6]。最低収入を保証されていると、様々なことにチャレンジできる。起業して、それが失敗しても生活だけは保証されるので、リスクを冒すことができる。

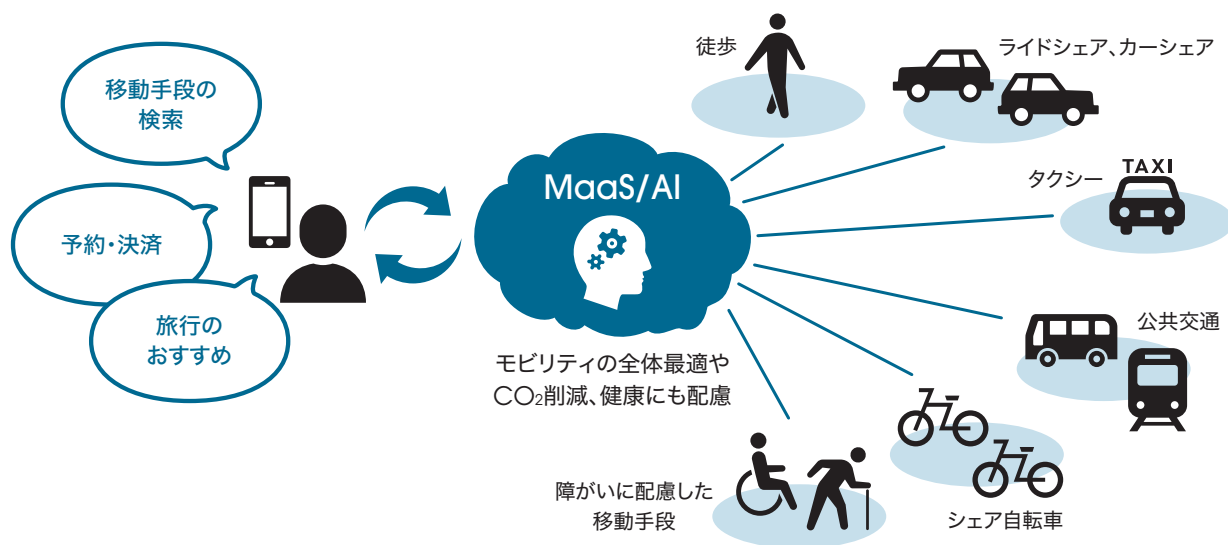
インターネットの出現によって、サービスを提供するためのコストがどんどん安くなっている。フィンテックもその一つである。IoTによるC2C基盤がUberやAirbnb等の「共有型経済」を実現し、そこでは消費者の既存資産(家や車)やインターネットを活用しているためサービスコストが限界的にはゼロに近づくような社会ができてくる。それによって今の資本主義から共有型の経済に変わることが必然になるという主張がある[7]。インターネットですべてのものを結び付ける(IoT)ことによって、シェアリングエコノミーが可能になる。経済の仕組みが根本的に変わるかもしれない。

### 2.6.5 モビリティ

シェアリングの流れは交通の世界にも変化を与えている。新しい公共交通への取組みは北欧が先進的である。ユーザーに移動というサービスを提供するMobility as a Service (MaaS) と呼ばれる概念が提唱されている。コンピューターサービスのクラウド化と同様、MaaSとは、提供される移動サービスとそれを提供するハードウェアを独立にする仕組みのことと理解すれば良い。UberやDiDiの配車システムが有名であるが、これらは現在のタクシーシステムを前提とした仕組みであり、MaaSとは呼び難い。

MaaS導入の先進事例として有名なヘルシンキ市では、タクシー、バス、鉄道、飛行機などの輸送機関が互いに協力、情報交換するための情報システムと仕組み作りを行っている。その象徴的なアプリとして、ベンチャー企業MaaS Global社がWhimと呼ばれるスマートフォン向けアプリを提供している(図2-6-3)。

■ 図2-6-3 MaaS(Mobility as a Service)のイメージ



Whimにメンバー登録しておく、利用者はタクシー、レンタカー、公共交通、シェア自転車などに自由にアクセスできるようになる。Whimは、利用者の利用傾向を学習し、予定表と連動してその日の目的地への最も効率的な移動法を提案してくる。多様な移動サービスを単一のインターフェースで効率よく利用可能としている。

トヨタもe-Palette Conceptを提唱しているが、これはコマツのIoTモデル(全世界のコマツの重機をインターネットで管理する)と似たもので自動車を売のではなく、車同士を情報的につないだ「コネクテッドカー」を使ったサービスモデルである。

公立はこだて未来大学発ベンチャーの未来シェアはさらに別のMaaSを提唱している。個人にとって移動は手段であって目的ではない。目的は医療であったり、飲食であったり、観光であったりする。インターネット上に様々なサービスが載っているがごとく、移動プラットフォーム上にこれらのサービスを載せることを目指して、AIによる移動の最適化を行っている。

### 2.6.6 医療

IBMのWatsonの医療応用が実用化されている。新薬、症例、治療法などの最新知識をアップデートしておくのは人間よりAIのほうが得意である。Deep LearningによるX線画像の分析も人間の診断医より高い精度で行える例も出てきた。

da Vinciなど、手術用の遠隔操作ハンド(手術支援ロボット)も開発されているが、これらもどんどん知的になっていくであろう。例えば操作する医師のミスにより対象外の血管を切るのを防ぐなどの制御ができよう。自動手術も視野に入る。

新薬開発において、標的となるタンパク質の同定や、新薬候補の化合物のスクリーニング及び臨床試験などの分野にAIを使う企業も増えてきている。

看護の世界でもAIやロボットが活躍するに違いない。精神面をケアするのは人間が主となろうが、見守りにはAIが使える。また、意思表示の困難な患者とのコミュニケーション支援もAIでできる。入浴介助や下の世話などはロボットの出番である。

理化学研究所では健康脆弱化予知予防コンソーシアムを立ち上げ、ビッグデータからの健康脆弱化予知(「未病」よりさらに前の段階で発見)することを試みているが、医療データの入手は個人情報保護法などがあり、容易ではない。東北大学はビッグデータメディスンセンターを設立し、大量の医療データの蓄積と利用による未来型医療を目指している。

### 2.6.7 教育

今後はCBT(Computer-Based Test)が普及すると考えられる。問題の提示と採点をコンピューターが行うというだけでなく、受験者のレベルに合った問題を提示することにより、従来より正確に受験者のレベルを判定できるようになる。IRT(Item Response Theory)を使い、解答を見ながら受験者のレベルに近い問題に絞り込んでいく。

現状では無理だが小論文の採点もできるようになるだろう。

IRTは教育にも使える。知識の全体系をAIが把握していて、学習者に最適なレベルの知識を与えていくのだ。昔あった「プログラム学習」のAI版である。

AIなどの技術発展が加速していくと専門知識もどんどん更新される。あるいは新しい専門分野ができたり、逆に特定の専門分野がなくなってしまうということも起こるだろう。大学で身につけた専門知識で一生仕事ができる時代はすでに過去のものである。

一方でAIが専門知識を学ぶことを考えると、人間が身につけなければならない知識とAIに頼れば良い知識の区別が必要になる。基本的には専門知識は(適宜必要な知識を人間に教育することを含め)AIに任せ、人間はリベラルアーツを身につけることになろう。リベラルアーツとはギリシャ・ローマ時代に端を発し、市民であるために必要な教養として教えられた文法、修辞学、論理学、算術、幾何学、天文学、音楽の7科目のことである。参考文献[8]は現代の教養(リベラルアーツ)として7科目(宗教、宇宙、人類の旅路、人間と病気、経済学、歴史、日本と日本人)を提案している。現在はこれに「情報技術」を加えねばならない(7つに収めるなら、最後の「日本と日本人」を歴史に入れればいい)。このように必要なリベラルアーツは時代とともに変わるが、重要なのはこれらが考え方の枠組みを作り上げるものであって、個別の専門知識ではないということである。従ってこれらの科目は(AIではなく)人間が教えなければならない。

#### ◆参考文献

- [ 1 ] 坂井豊貴『「決め方」の経済学 ―「みんなの意見のまとめ方」を科学する』ダイヤモンド社, 2016.
- [ 2 ] 富山和彦『AI経営で会社は甦る』文藝春秋, 2017.
- [ 3 ] Harvard Business Review 2018年7月号「特集:アジャイル人事 俊敏な組織に進化する」ダイヤモンド社, 2018.
- [ 4 ] ダイアン・マルケイ(門脇弘典訳)『ギグ・エコノミー 人生100年時代を幸せに暮らす最強の働き方』日経BP社, 2017.
- [ 5 ] トマ・ピケティ(山形浩生、守岡桜、森本正史 訳)『21世紀の資本』みすず書房, 2014.
- [ 6 ] 井上智洋『AI時代の新・ベーシックインカム論』光文社新書, 2018.
- [ 7 ] ジェレミー・リフキン(柴田裕之訳)『限界費用ゼロ社会 <モノのインターネット>と共有型経済の台頭』NHK出版, 2015.
- [ 8 ] 池上彰『おとなの教養 私たちはどこから来て、どこへ行くのか?』NHK出版新書, 2014.

## 2.7 ▷ AI人材の育成

### 2.7.1 AI人材育成の全体イメージ

経済産業省の2016年度調査「IT人材の最新動向と将来推計に関する調査結果」では、今後特に大幅な市場拡大が予想される「ビッグデータ」、「IoT」、「人工知能」を担う人材（先端IT人材）については2020年に約4.8万人不足するという予測が出されている。IPAが2017年度に実施したアンケートでも、AIベンダー側の「AIビジネス推進上の課題」として「自社にAIに関する人材が不足している」がトップとなっている（第5章参照）。一般的に人手不足である状況に加え、新しい技術・知識を要することから、AI開発現場におけるAI関連人材の獲得は急務と想定される。

このような状況に対して2017年6月に閣議決定された「未来投資戦略2017」の中でも、Society 5.0に向けた横割課題の一つとして「教育・人材力の抜本強化」が挙げられており、AIやデータ活用を含む「IT力」により目指すべき社会像を実現するための項目が示されている。また、2018年6月に閣議決定された「未来投資戦略2018」では「AI時代に対応した人材育成と最適活用」として、学校教育におけるAI・データ活用を含むIT人材の育成などの施策がさらに提言されている（表2-7-1）。

■表2-7-1 未来投資戦略2017/2018における人材関連項目

戦略	項目	概要
未来投資戦略 2017	「何を学ぶべきか」の羅針盤の提示	IT人材需給を把握する仕組みの構築、ITスキル標準改訂等
	産学官連携による実践的教育	大学協議体、専修学校等との産学連携の取組み、官民コンソーシアムの設立等
	大学の数理・データサイエンス教育の強化、工学教育改革等	数理・データサイエンス教育センター整備、小学校でのプログラミング教育等
	誰もが学び直しできる社会	第四次産業革命スキル習得講座認定制度創設、学び直しの支援策等
未来投資戦略 2018	AI時代に対応した人材育成と最適活用	小学校でのプログラミング教育に向けた環境整備等
		大学入学共通テストでの「情報」の必修科目化、大学での数理・データサイエンス一般教養化等
		AI人材の育成のための分野横断的で実践的な人材育成を行うための「学位プログラム」の実現等
		デジタル・トランスフォーメーション、AI・データリカレント教育、AI人材の高待遇化等
		副業・兼業を通じたキャリア形成を促進するための実効性のある労働時間管理等の在り方の検討等

出典：未来投資戦略2017、2018より作成

また、2018年6月に閣議決定された「統合イノベーション戦略」では、「AI技術」の目指すべき将来として、

- ・ これからの「読み・書き・そろばん」であるAI技術を使いこなすITリテラシーを誰もが持ち、ヒューマンフレンドリーなAI技術を活用することで、ニーズに合った物・サービスの提供、病気になるないヘルスケア、自由で安全な移動等を実現
- ・ サイバーセキュリティが確保され、AI技術の社会受容が進み、産業から生活まで様々な分野で活用されることで、質の高い新たな雇用やサービスを創出

が挙げられており、そのための「人材基盤の確立」として、以下の目標が掲げられている。

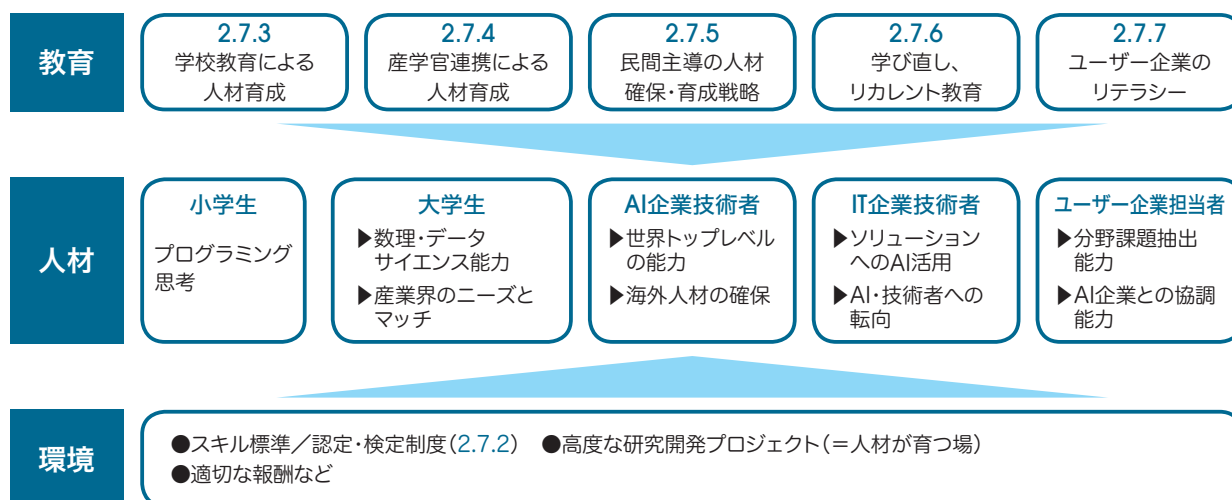
- ・2025年までに先端IT人材を年数万人規模、IT人材を年数十万人規模で育成・採用できる体制を確立(関係府省庁の施策での育成規模を2018年度中に設定)
- ・2032年までに初等中等教育を終えたすべての生徒がITリテラシーを獲得

上記を整理すると、AI・データ関連人材の育成には、スキル標準の整備によりスキルセットやキャリアパスを明確にしつつ、

- ・初等中等から大学までの課程における一般教養としての位置づけ
- ・産学官連携による産業ニーズを踏まえた学校教育
- ・高度な研究開発プロジェクトによる場の提供
- ・社会人のAI・データ人材への学び直しやリカレント教育

などの育成システムの充実を図る必要がある。また、産業へのAI導入を推進するためには、産業・自社特有の課題を洗い出したり、データを収集・活用したりするためのユーザー企業担当者のAIリテラシーも重要と考えられる。図2-7-1に全体イメージを示す。

■図2-7-1 AI・データ関連人材育成の全体イメージ



本節では、2.7.2で「スキル標準／認定・検定制度」、2.7.3で「学校教育による人材育成」、2.7.4で「産学官連携による人材育成」、2.7.5で「民間主導の人材確保・育成戦略」、2.7.6で「学び直し、リカレント教育」、2.7.7で「ユーザー企業のリテラシー」について整理する。

## 2.7.2 スキル標準／認定・検定制度

### (1) IPA 「ITSS+ (プラス)」

IPAでは、各種IT関連サービスの提供に必要とされる能力を明確化・体系化した指標であり、産学におけるITサービス・プロフェッショナルの教育・訓練等に有用な「ものさし」(共通枠組み)を提供しようとするものとして2002年に経済産業省が策定した「ITスキル標準(ITSS)」の普及・活用促進を図っている。

2017年4月には、第4次産業革命に向けて求められる新たな領域の“学び直し”の指針として「ITSS+(プラス)」を策定している。「ITSS+」は「データサイエンス領域」及び「セキュリティ領域」を対象とした形で2017年4月に公開、2018年4月には「IoTソリューション領域」及び「アジャイル領域」

を追加している。ITSS+のうち、AIに関連する「データサイエンス領域」は、企業等の業務において大量データを分析し、その分析結果を活用するための一連のタスクとそのために習得しておくべきスキルを取りまとめたものであり、タスクはIPAと「一般社団法人データサイエンティスト協会 スキル委員会(委員長：安宅和人 ヤフー株式会社 CSO)」の協業で策定している。

## (2) 一般社団法人日本ディープラーニング協会「G検定／E資格」

日本ディープラーニング協会(2017年6月設立)は、「ディープラーニングの基礎知識を有し、適切な活動方針を決定して事業応用する能力を持つ人材」すなわちディープラーニングを含むAIの知識・リテラシーを持ったビジネスサイドの人材と、「ディープラーニングの理論を理解し、適切な手法を選択して実装する能力を持つ人材」すなわちエンジニア人材の双方の育成が必要との考えから、それぞれについて「G(ジェネラリスト)検定」「E(エンジニア)資格」なる試験を開催している。G検定の第1回試験結果は2018年1月に発表され、1,448人の受験者のうち、823人が合格している(合格率56.8%)。G検定とE資格にはそれぞれシラバスが定義されており、Web上で以下のとおり公開されている(表2-7-2、表2-7-3)。

■表2-7-2 G検定のシラバス(2018)

項目	概要
人工知能(AI)とは	人工知能の定義
人工知能をめぐる動向	探索・推論、知識表現、機械学習、深層学習
人工知能分野の問題	トイプロブレム、フレーム問題、弱いAI、強いAI、身体性、シンボルグラウンディング問題、特徴量設計、チューリングテスト、シンギュラリティ
機械学習の具体的な手法	代表的な手法、データの扱い、応用
ディープラーニングの概要	<ul style="list-style-type: none"> <li>ニューラルネットワークとディープラーニング、既存のニューラルネットワークにおける問題、ディープラーニングのアプローチ、CPUとGPU</li> <li>ディープラーニングにおけるデータ量</li> </ul>
ディープラーニングの手法	<ul style="list-style-type: none"> <li>活性化関数、学習率の最適化、更なるテクニック、CNN、RNN</li> <li>深層強化学習、深層生成モデル</li> </ul>
ディープラーニングの研究分野	画像認識、自然言語処理、音声処理、ロボティクス(強化学習)、マルチモーダル
ディープラーニングの応用に向けて	産業への応用、法律、倫理、現行の議論

■表2-7-3 E資格のシラバス(2018)

項目	概要
応用数学	線形代数、確率・統計
機械学習	機械学習の基礎、実用的な方法論
深層学習	順伝播型ネットワーク、深層モデルのための正則化、深層モデルのための最適化、畳み込みネットワーク、回帰結合型ニューラルネットワークと再帰的ネットワーク、自己符号器、生成モデル、強化学習

出典：一般社団法人日本ディープラーニング協会 Webページより

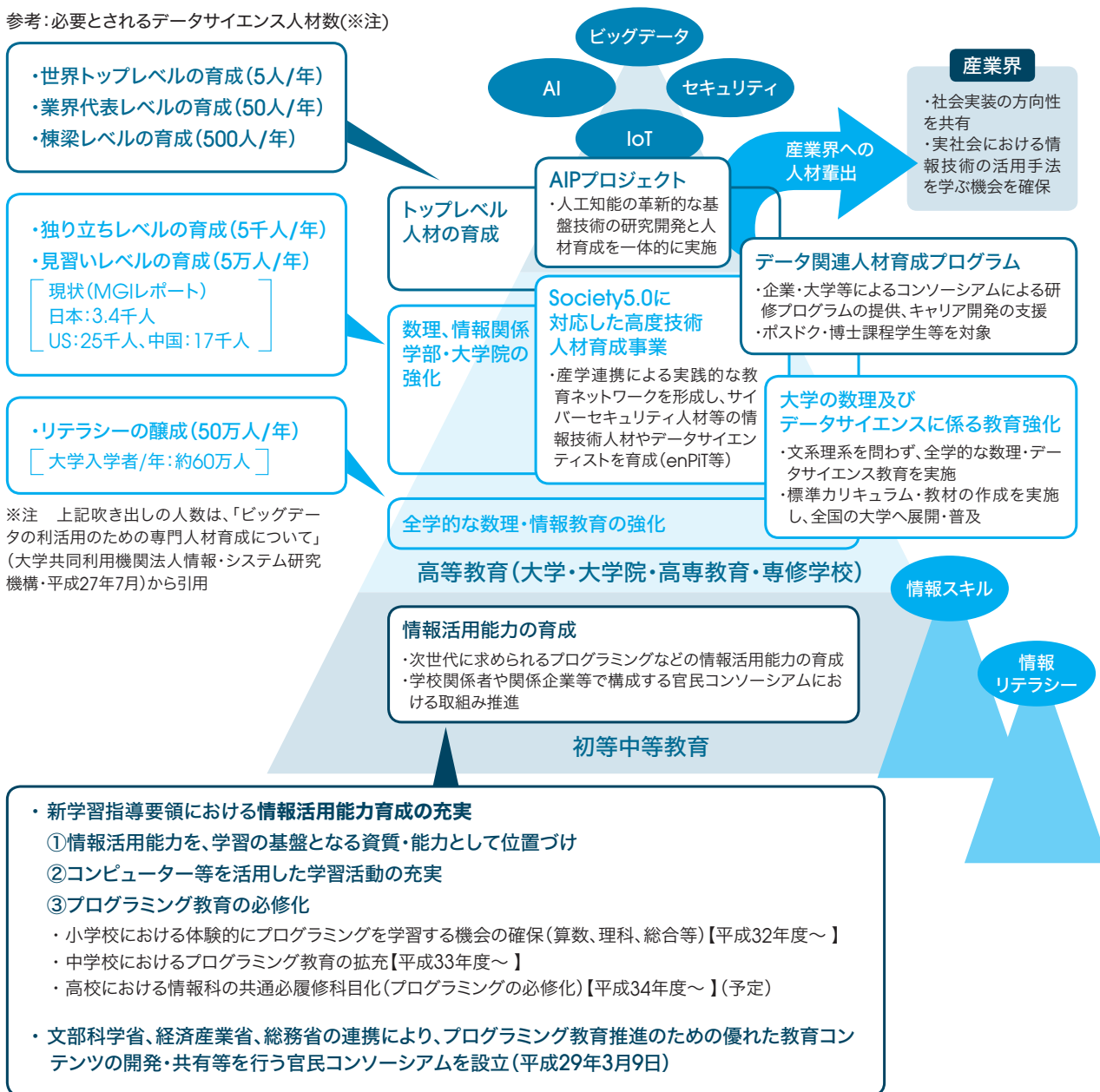
E資格については、協会のシラバスに沿った講義とプログラミング演習を行う「JDLA認定プログラム」の受講が受験申込条件となっている。この「JDLA認定プログラム」の枠組みは、プログラミングスキルやモデルチューニングのスキルなど短時間のペーパーテストだけでは測れないスキルの育成と、ディープラーニングを学べる場を増やすことを目的としている。

## 2.7.3 学校教育による人材育成

文部科学省では小学校からプログラム教育や大学での数理及びデータサイエンスなど、学校教育におけるIT活用能力の育成に注力している(図2-7-2)。

■図2-7-2 Society 5.0の実現に向けた総合的な人材育成の概念図

参考:必要とされるデータサイエンス人材数(※注)



※注 上記吹き出しの人数は、「ビッグデータの利活用のための専門人材育成について」(大学共同利用機関法人情報・システム研究機構・平成27年7月)から引用

出典:総合科学技術イノベーション総合会議「Society 5.0実現に向けた戦略的重要課題について」文部科学省資料

### (1) 情報活用能力の育成

小学校段階でのプログラミング教育必修化(2020年度～)に向け、学校現場での楽しみながら学べるデジタル教材の活用・評価とさらなる改善等の産業界と教育現場が連携した取組みを2018年度秋から開始、2019年度から本格展開する。

### (2) 大学の数理・データサイエンスに係る教育強化

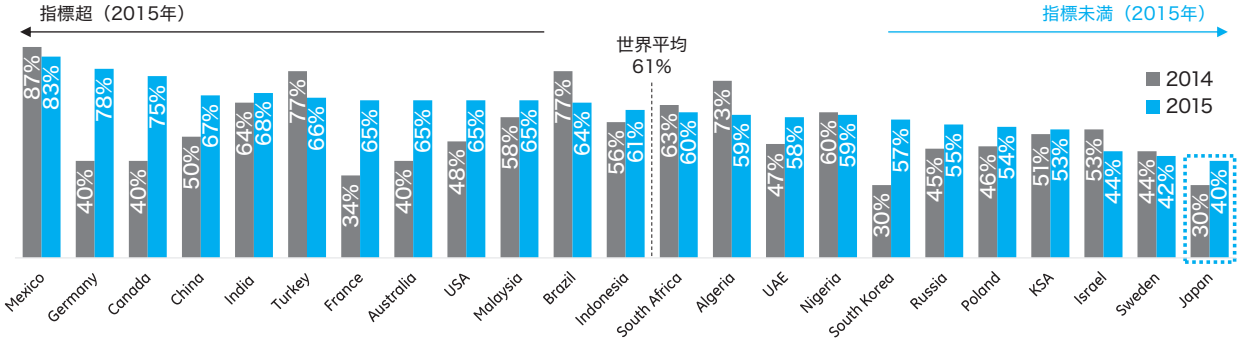
文部科学省では2016年12月、数理やデータサイエンス教育の強化に関する懇談会において大学の数理・データサイエンス教育強化方策を公表している(図2-7-3)。

■ 図2-7-3 大学の数理・データサイエンスに係る教育強化(概要)

- 現状**
- 膨大なデータが溢れる時代において、諸外国と比較すると企業では意思決定におけるデータとアナリティクスの活用に遅れをとっている状況
  - 世界に先駆けた「超スマート社会」の実現(Society5.0)に向けて、我が国の産業活動を活性化させるために必要な数理・データサイエンスの基礎的素養を持ち、課題解決や価値創出につなげられる人材育成が必要不可欠

○我が国の企業幹部におけるデータの分析・活用の戦略的価値への認識は、世界の主要国に比べて非常に低い

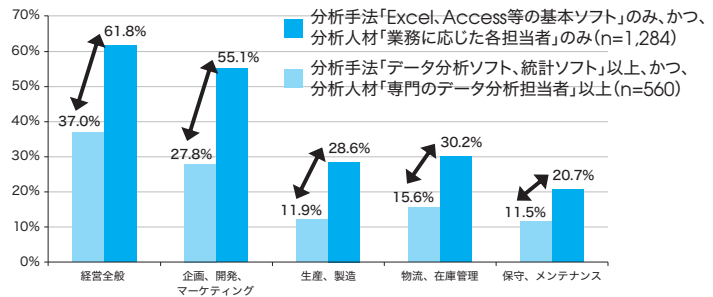
●企業幹部におけるデータとアナリティクスを用いた意思決定割合



出典:「GEグローバルイノベーション・バロメーター2016」世界の経営層の意識調査

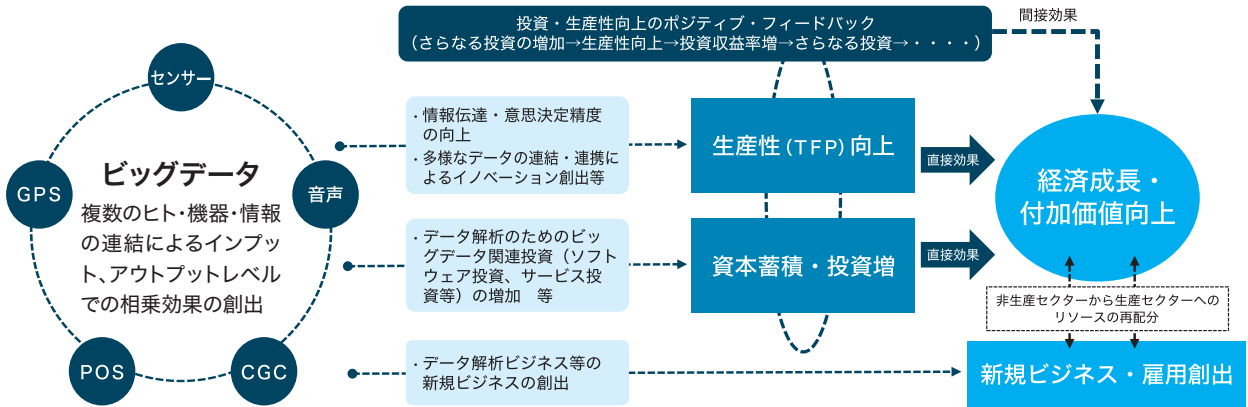
○数理的思考やデータ分析・活用能力を持つ人材が戦略的にデータを扱うことによる経営等への効果は大きい

●分析手法・分析人材の違いによる効果割合



出典:総務省「ビッグデータの流通量の推計及びビッグデータの活用実態に関する調査研究」(平成27年)

●データの流通・蓄積・活用による産業活動の活性化

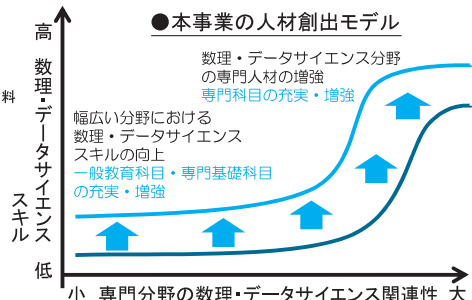
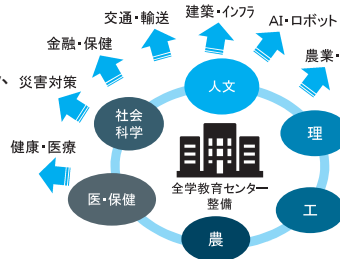


出典:総務省「情報流通・蓄積量の計測手法の検討に係る調査研究」(平成25年)

専門分野の枠を超えた全学的な数理・データサイエンス教育機能を有するセンターを整備し、専門人材の専門性強化と他分野への応用展開の双方を実現し、相乗効果を創出

実現に向けたシナリオ

- 文系理系を問わず、全学的な数理・データサイエンス教育を実施
- 医療、金融、法律等の様々な学問分野へ応用展開し、社会的課題解決や新たな価値創出を実現
- 実践的な教育内容・方法の採用
  - ・企業から提供された実データ等のケース教材の活用
  - ・グループワークを取り入れたPBLや実務家による講義等の実践的な教育方法の採用
- 標準カリキュラム・教材の作成を実施し、全国の大学へ展開・普及



出典:文部科学省/数理及びデータサイエンス教育の強化に関する懇談会「大学の数理・データサイエンス教育強化方策について」(2016年12月)



これを受けて、2017年12月、北海道大学、東京大学、滋賀大学、京都大学、大阪大学、九州大学の6校が拠点校として選定されている。また、当該6校により、数理・データサイエンス教育強化拠点コンソーシアムが設立されており、国際的なモデルとなる標準カリキュラム・教材の作成、他大学への普及方策などの検討が行われている<sup>\*64</sup>。

## 2.7.4 産学官連携による人材育成

ここでは、産業界のニーズに応じた人材を大学で育成したり、産業界の即戦力を大学の場を活用して育成したりする文部科学省の施策を示す。

### (1) 大学協議体と産業界による意見交換

大学協議体は、産業界のニーズを継続的に把握しつつ、産業界の代表との実務レベルでの情報共有等を行うことを目的として、「未来投資会議2017」にもとづいて国公私立大学の学部長等により組織された。具体的な取組み内容は以下のとおりである。

- ・教育機関側と産業界側それぞれに対する要望について、大学協議体と産業界の実務レベルによる意見交換
- ・産学協働による教育プログラムの構築・実施や調査等にもとづく政策提言の取りまとめ

2018年3月には、「産学連携による科学技術人材育成に関する大学協議体と産業界による意見交換(第1回)」が開催され、「産業界のニーズと高等教育のマッチング方策」、「産学連携教育の推進」、「リカレント教育(技術者の継続教育)の推進」といった論点で議論が行われている。資料の中では、今後不足が見込まれるIT分野としてデータ・AI人材が挙げられる部分もあるが、現段階では基本的には枠組みの検討が中心である。この枠組みを通じて、AI・データ関連業界のニーズが大学側に伝われば、リカレント教育を含めた人材育成の推進が期待される。

### (2) 専修学校による地域産業中核的人材養成事業

地域産業の発展を支える中核的な人材養成機関としての専修学校の役割の充実等を図るため、社会人向けの教育プログラムや特色ある教育カリキュラムの開発、効果的な産学連携教育の実施のためのガイドラインの作成、分野に応じた中長期的な人材育成に向けた協議体制の構築に係る事業の委託を行う制度である。2017年度の採択テーマの中には、ビッグデータやAIも一部含まれている(2018年度事業は公募中)。

### (3) Society 5.0に対応した高度技術人材育成事業

産学連携による実践的な教育ネットワークを形成し、産業界のニーズに応じた人材を育成する取組みを支援する事業であり、2018年度は以下の2事業を公募。

- 超スマート社会の実現に向けたデータサイエンティスト育成事業
- 科学技術の社会実装教育エコシステム拠点の形成事業

前者では、産学官による実践的な教育ネットワークを構築し、文系理系を問わず様々な分野へデータサイエンスの応用展開を図り、それぞれの分野でデータから価値を創出し、ビジネス課題や社会課題に答えを出す人材(データサイエンティスト)を育成する大学を支援する。

※64 数理・データサイエンス教育強化拠点コンソーシアム<<http://www.mi.u-tokyo.ac.jp/consortium/index.html>>

#### (4) データ関連人材育成プログラム

第4次産業革命を勝ち抜くうえで求められるAI、IoT、ビッグデータ、セキュリティ等を高度に駆使する人材(高度データ関連人材)について、発掘・育成・活躍促進を一貫して行う企業や大学等における取組みを支援することで、データ利活用社会のエコシステム構築への貢献を目指す。

- ・研究活動を通じて高度なデータの扱いに親しんだ博士課程学生や博士号取得者等を対象
- ・企業や大学等が人材の発掘・育成・活躍促進を目的としたコンソーシアムを形成
- ・インターンシップやPBL(課題解決型学習)等の実践的な研修プログラムを開発・実施

### 2.7.5 民間主導の人材確保・育成戦略

#### (1) 人材確保戦略

産業界におけるAI人材関連動向としては、社内にAI研究開発部門を設けて人材を集約したり、複数企業の共同で人材を含めたリソースを連携させたり、AI研究開発拠点を海外に設置することで国外のAI人材を活用したりする戦略が見られる。

#### ○国内の動向

『AI白書2017』では、大手企業におけるAI開発に関する人材確保や組織体制強化に関する動向を紹介しているが、この流れは現在も継続している。以下に例を示す。

- ・トヨタ自動車株式会社、アイシン精機株式会社、株式会社デンソーは2018年3月、知能化ソフトウェアの研究から開発を一気通貫で担う新会社「Toyota Research Institute Advanced Development」を東京に設立することを発表。1,000名規模の体制とすることを目標とし、アクセス面や採用面で魅力のある新たなロケーションを選定。「新会社の狙い」の一つに、「国内外トップ人材採用により開発力を強化しつつ、トヨタグループ内の知能化人材を育成」が含まれている<sup>\*65</sup>。
- ・日立製作所はデータサイエンティストをグローバルに育成・強化するため、スキル要件と育成プログラムを整備するとともに、プロフェッショナル・コミュニティを立ち上げ、国内外のグループ会社におけるデータサイエンティストを現在の700名から2021年度までに3,000名にすることを目標に増強すると発表<sup>\*66</sup>。
- ・NECは2018年4月26日、AIを活用した新たなデータ分析事業の展開として、データ分析ツールの開発及び販売、コンサルティングを事業とする新会社を米国シリコンバレーに設立すると発表。シリコンバレーの高度人材の雇用を狙うとのこと<sup>\*67</sup>。
- ・株式会社安川電機は2018年3月1日、AIソリューション開発の子会社を設立。同時にAIベンチャー企業のカロスコンパス社と戦略的提携も公表しており、外部のAIリソースを活用<sup>\*68</sup>。
- ・伊藤忠テクノソリューションズ株式会社、株式会社グリッド、TIS株式会社、富士通株式会社など11社は2017年12月13日、「AI ビジネス推進コンソーシアム」を設立。ノウハウを相互に共有、連携を図り、ビジネス利用を促進することを目的としており、AIエンジニアの人材育成にも力を入れる<sup>\*69</sup>。

※65 トヨタ自動車ニュースリリース<<https://newsroom.toyota.co.jp/jp/corporate/21313423.html>>

※66 日立製作所ニュースリリース<<http://www.hitachi.co.jp/New/cnews/month/2018/06/0621.html>>

※67 NECプレスリリース<[https://jpn.nec.com/press/201804/20180426\\_01.html](https://jpn.nec.com/press/201804/20180426_01.html)>

※68 安川電機のお知らせ<<https://www.yaskawa.co.jp/newsrelease/news/45421>>

## ○国外の動向

AI開発に関する人材確保や組織体制強化については海外企業が先行している印象があるが、この1年においても国内外に研究所を設置し、AI研究者を確保しようという動きがみられる。

- ・米IBMとMITは2017年9月7日、「MIT-IBM Watson AIラボ」を共同で設立し、今後10年間で2億4000万ドル(約272億円)を投資すると発表。およそ100人のAIを専門とする研究者や学生、専門家が参加の見込み<sup>\*70</sup>。
- ・米Microsoftは2017年7月12日、AI専門研究ラボ「Microsoft Research AI」の設立を発表。同社は前年9月にも5000人規模の「Microsoft AI and Research Group」を設立しており、本ラボはより汎用的な学習システムの構築を目的とする。100人以上の研究者が取り組む見込み<sup>\*71</sup>。
- ・米Googleは2017年12月、北京にAIセンター「the Google AI China Center」を設立。中国のトップエンジニアを雇用し、AI研究チームを構成する模様<sup>\*72</sup>。
- ・中国のアリババ・グループ・ホールディングは、AIやIoT、量子コンピューティング等の研究のために、全世界に7カ所の研究所を設置し、研究員を100人採用すると電子メールを通じて発表した。研究開発費を今後3年で150億ドル(約1兆6,900億円)とする<sup>\*73</sup>。
- ・韓国のサムスン電子が2017年の韓国ソウル、米シリコンバレーに続き、2018年5月に英ケンブリッジ、加トロント、露モスクワにAIセンターを設置、世界のAIの専門家を確保することによって、高度なAI研究者を2020年までに1,000人に拡大する計画<sup>\*74</sup>。

## (2) 人材育成戦略

### ・先端人工知能学教育寄付講座

ディープラーニングを含む先端AI技術と、その理論基盤に関する体系的教育プログラムの構築・実施による人材育成を目的として、2016年6月、東京大学に開設された。2021年5月までの5年間の活動を予定している。トヨタ自動車、ドワンゴ、オムロン、パナソニック、野村総合研究所、DeNA、みずほフィナンシャルグループ、三菱重工業の計8社から合計9億円の寄付により実現した。約50人の大学院生に対して先端AI技術に関する高度な教育を行っている。

### ・AIを活用した素材開発

東レ、昭和電工、積水化学、富士フイルム、三井化学など素材大手が、東京大学や一般社団法人日本ディープラーニング協会と連携し、ビッグデータや機械学習を素材開発に活用できる人材の育成事業を始める。5年間で1,500人の育成を計画している。

※69 AIビジネス推進コンソーシアム<<http://Albpc.org/>>

※70 IBMニュースリリース<<https://www-03.ibm.com/press/jp/ja/pressrelease/53130.wss>>,IBM THINK Business<<https://www.ibm.com/think/jp-ja/watson/pressrelease/mit/>>

※71 アイティメディアリリース<<http://www.itmedia.co.jp/news/articles/1707/13/news061.html>>,Microsoft Research AI<<https://www.microsoft.com/en-us/research/lab/microsoft-research-AI/>>

※72 Reuters Business News<<https://www.reuters.com/article/us-google-china/google-launching-artificial-intelligence-research-center-in-china-idUSKBN1E70A3>>,Google Careers<<https://careers.google.com/stories/opening-the-google-AI-center-in-china/>>

※73 Bloomberg News<<https://www.bloomberg.co.jp/news/articles/2017-10-11/OXN1C56S972901>>

※74 Samsung Newsroom<<https://news.samsung.com/global/samsung-opens-global-AI-centers-in-the-u-k-canada-and-russia>>

## 2.7.6 学び直し、リカレント教育

AI人材の不足を補うためには、既存の社会人のスキルチェンジも重要である。

### (1) リカレント教育

経済産業省では、「人生100年時代の社会人基礎力」として、社会人が学び直しを通じたアップデートや新たなスキルの獲得を継続することが必要不可欠として、基礎能力(OS)を身につけつつ業界特性等に応じた能力(アプリ)を新規獲得及びアップデートしていく必要性を提言している(図2-7-4)。

■ 図2-7-4 「人生100年時代の社会人基礎力」を伸ばすためのリカレント教育

- ◎平成30年度に、中小企業が中核人材に①求める能力(キャリア・オーナーシップ、社会人基礎力)や②求めるスキル(専門スキル)に関する講座(コンテンツ)を、従業員の業務内容やステージ(役職)に応じて作成。
- ◎これまでの一方通行型の座学のみならず、Ed-Techを活用した多様な学びのスタイルで提供する。
- ◎加えて、受講前後のスキル診断と、受講後の反復学習を併せて行うことにより定着度を高めるとともに、各講座の受講履歴等を一元的に管理することで、理解度を可視化する。

	【コンテンツ(何を学ぶのか)】	【提供スタイル(どのように学ぶのか)】
OS (能力)	<b>1. キャリア・オーナーシップ</b> ○スキルの棚卸し・キャリアプランの見直し等  <b>2. 社会人基礎力</b> ○問題解決型思考の獲得 ○リーダーシップの構築/自律性の獲得 (主体的な意思決定・行動)等	<b>1. 双方向簡易型スタイル【ウェブ型】</b> ※時間がない社会人でも「すきま時間」を活用した学びが可能 ○受講時間：いつでも ○受講場所：どこでも ○受講方法：PC・スマホ・タブレット  <b>2. 双方向ライブ配信スタイル【ライブ型】</b> ※オンライン機能を活用し、双方向型の学びが場所を選ばず可能 ○受講時間：ライブ日時は固定・録画はいつでも・どこからでも ○受講場所：どこでも ○受講方法：PC・スマホ・タブレット
アプリ (スキル)	<b>3. 専門スキル</b> ○IT活用(リテラシー)講座 ○中小企業からのニーズの高い「営業・販売・サービス」や「生産現場」等に関するマネジメント講座等	<b>3. ワークショップスタイル【対面型】</b> ※一度に集中して学ぶことが可能 ○受講時間：固定 ○受講場所：指定された教室等 ○受講方法：集合・座学

出典：経済産業省 我が国産業における人材力強化に向けた研究会「人生100年時代の社会人基礎力」と「リカレント教育」について

### (2) 第四次産業革命スキル習得講座認定制度

経済産業省は2017年7月に「第四次産業革命スキル習得講座認定制度」を創設している。同制度は、IT・データを中心とした将来の成長が強く見込まれ、雇用創出に貢献する分野において、社会人が高度な専門性を身につけてキャリアアップを図る、専門的・実践的な教育訓練講座を経済産業大臣が認定する制度である。厚生労働省が定める一定の要件を満たし、厚生労働大臣の指定を受けたものは、「専門実践教育訓練給付」の対象となる。本制度で対象となる分野は、AI、IoT、データサイエンス、クラウドの他、セキュリティなども含む。表2-7-4に、AI・データサイエンス関連の認定講座を示す。

■表2-7-4 第四次産業革命スキル習得講座認定制度 認定講座一覧(AI・データサイエンス関連)

AI関連(講座名は仮称を含む)		
第一回	株式会社チェンジ	「AI活用コンサルタント」育成トレーニング～Aler 育成プログラム～
	株式会社ウチダ人材開発センタ	AI活用講座
	日本マイクロソフト株式会社	ディープラーニングハンズオンセミナー
	株式会社富士通ラーニングメディア	Fujitsu Digital Business College/AI・データ分析を活用するイノベーター
第二回	株式会社北海道ソフトウェア技術開発機構	AIエンジニア講座
	AI TOKYO LAB株式会社	AIエンジニア講座
	学校法人金沢工業大学	AIビジネスアドミニストレータ養成講座/AI技術アドミニストレータ養成講座/AIジェネラリスト養成講座/AIエヴァンジェリスト養成講座
	エッジコンサルティング株式会社	機械学習講座
	株式会社エナジャイズ	デザインシンキング講座/AIエンジニア講座
	株式会社DIVE INTO CODE	エキスパートAIコース
	学校法人先端教育機構	AIエンジニア講座
	トレノケート株式会社	ディープラーニング ハンズオンセミナー
データサイエンス関連(講座名は仮称を含む)		
第一回	株式会社efftax	データ分析教育講座白・茶・黒帯編
	株式会社ブレインパッド	データサイエンティスト入門研修
	株式会社ブレインパッド	データサイエンティスト入門研修(アドバンスド)
	株式会社アイ・ラーニング	データサイエンティスト育成講座
	株式会社チェンジ	データサイエンティスト養成コース
	株式会社日立インフォメーションアカデミー	データ活用技術者育成講座
	フューチャー株式会社	データサイエンティスト養成講座
第二回	一般財団法人リモート・センシング技術センター	リモートセンシングデータ解析技術者養成講座
	株式会社データミックス	データサイエンティスト育成コース パートタイムプログラム
	一般財団法人日本海事協会	海事データサイエンティスト育成講座【機器計測データの解析】
	株式会社アルベルト	データサイエンティスト養成講座(R言語 上級編)/データサイエンティスト養成講座(Python 上級編)

出典:経済産業省公表資料より

### (3) AI即戦力人材育成講座「AIデータフロンティアコース」(NEDO、東京大学/大阪大学)

国立研究開発法人新エネルギー・産業技術総合開発機構(NEDO)は2017年下期にAI即戦力人材の育成を目的に「AIデータフロンティアコース」を開講した。同コースは、社会人技術者及び研究者を主な対象としており、2018年度から本格スタート、2019年度まで実施される。

本コースでは、講義でAI知識を体系的に習得するとともに、製造現場や顧客行動等の実社会における様々なデータセットを用いた演習を通じてデータの構築方法や解析手法など、半年間でAI技術を身につけることができる。

本コースの拠点は大阪大学と東京大学であるが、これらの拠点を中心に受講者参加型のシンポジウムやワークショップを開催することで、多方面の人材の交流を図るとともに、関連技術を含めたAI分野の新たな技術シーズの発掘や技術の応用・発展に資する取組みを行い、当該技術を担う人材が育つという「好循環」を事業終了後も継続的に形成することを目指している(表2-7-5)。

■表2-7-5 NEDO/AIデータフロンティアコース事業内容

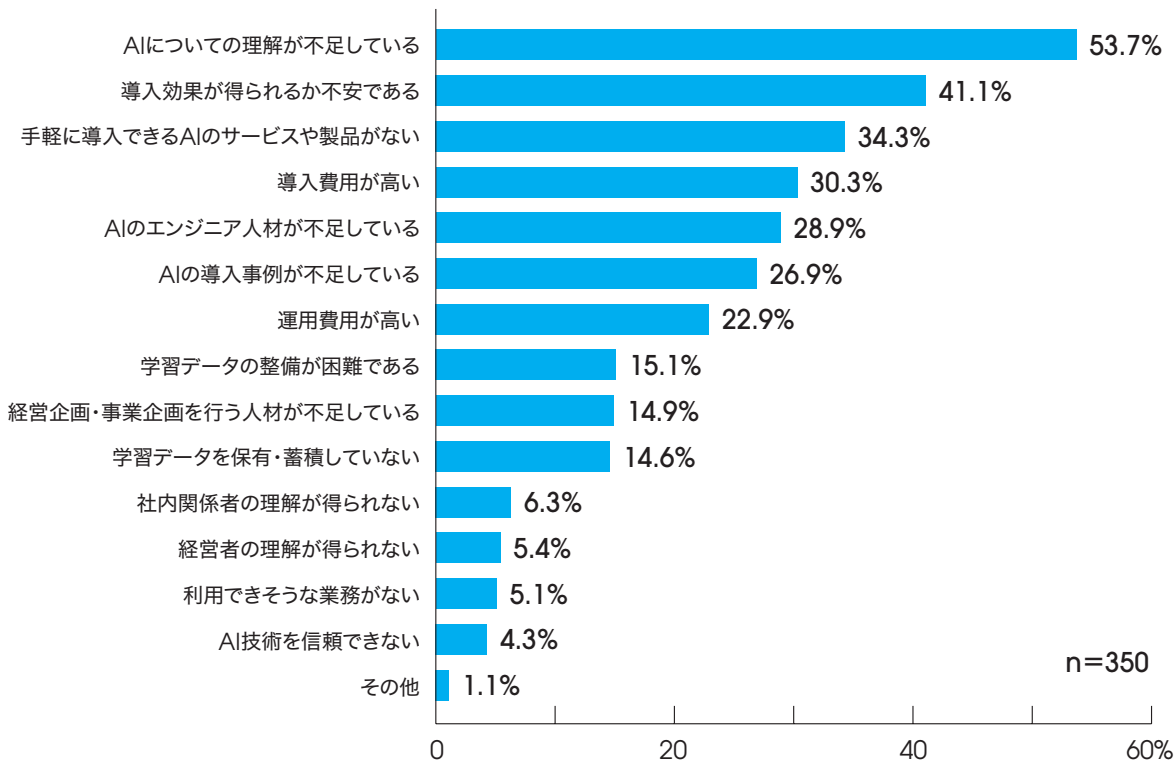
項目	概要
人的交流等の展開	コンピューターサイエンスの基礎学力テストと補習、AIに関するトップレベル講義、リアルコモンデータを扱う演習、演習終了時の能力評価を通じて、即戦力人材の育成を行う。
人材育成要請講座	人材育成拠点と受講者の所属企業、大学、関係機関等の人的交流を促進するため、受講者参加型のシンポジウムやワークショップを実施する。
周辺研究	効果的かつ継続的なAI人材育成を実施するため、実社会・実環境で収集し、教育目的において汎用的に活用できるデータセットの作成に取り組む。また、産業界のニーズを踏まえた最先端のAI知識と活用スキルを持つ人材を短期間で育成するための人材育成カリキュラムの開発も進める。

出典：新エネルギー・産業技術総合開発機構(NEDO)ニュースリリース「即戦力となるAI分野の人材を育成—NEDO特別講座を大阪大学と東京大学の2拠点で開講へ—」(2017年7月28日)に加筆

## 2.7.7 ユーザー企業のリテラシー

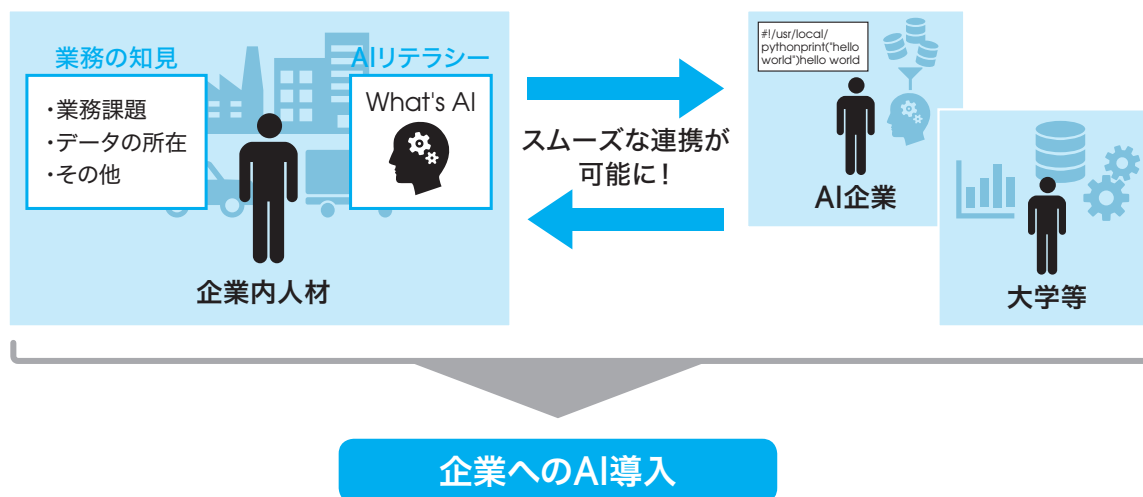
IPAが2017年度に実施したAI社会実装推進事業におけるアンケート調査では、ユーザー企業側の「AI導入検討上の課題」として「AIについての理解が不足している」が突出(図2-7-5)、AIベンダー側でも「ユーザー側のAIへの理解が不足している」が50%となっている。

■図2-7-5 ユーザー側におけるAI導入検討上の課題



製造業やサービス業などの一般企業がAIを導入する場合、必要なAI開発については専門企業への委託等で対応可能であるが、AIで解決可能な現場の課題を洗い出したり、データの有無の確認や収集・蓄積を行ったりすることは、当該産業の特性や社内業務を熟知している企業内人材が行うほうが効率的である。この企業内人材が、AI開発の委託・協業先の技術者とスムーズに取組みを進めるためには、AI専門用語やAI手法の概略などの「AIリテラシー」を身につける必要がある(図2-7-6)。

■ 図2-7-6 ユーザー企業内人材のAIリテラシーの必要性



しかしながら、社内業務を熟知している人材は企業内でも需要が高いため、AIリテラシーを身につけるための大学での学び直しやリカレント教育の時間確保は困難と想定される。大企業では社内に講師を呼んで研修を行うケースも見られるが、中堅中小企業においては容易ではない。平成30年6月に閣議決定された「統合イノベーション戦略」では、「AI技術」の目指すべき将来として「これからの「読み・書き・そろばん」であるAI技術を使いこなすITリテラシーを誰もが持ち、……」とあり、そのための初等中等教育でのITリテラシー教育が挙げられているが、現在のユーザー企業内人材のAIリテラシー習得も課題と想定される。

なお、一般社団法人日本ディープラーニング協会が実施するビジネスサイドのAI人材育成を目指す「G検定」(2.7.2参照)の合格者アンケートでは、「人工知能関連の知識を体系的に整理できた」、「AIベンダーやエンジニアとの会話が成り立つようになった」、「AI関連の製品や展示会の内容がわかるようになった」などの意見があったとのことであり、ユーザー企業内人材のAIリテラシー向上がAI導入推進につながることを期待される。2017年の合格者からは、実際に案件が立ち上げられた、ビジネスにつながったという声もあったという。ただし、「合格しても周囲でG検定を知っている人がいなかった」という意見も少なくないとのことであり、企業内人材の努力が適切に評価されるようAIに関する認定・検定制度の周知を図ることも必要と考えられる。