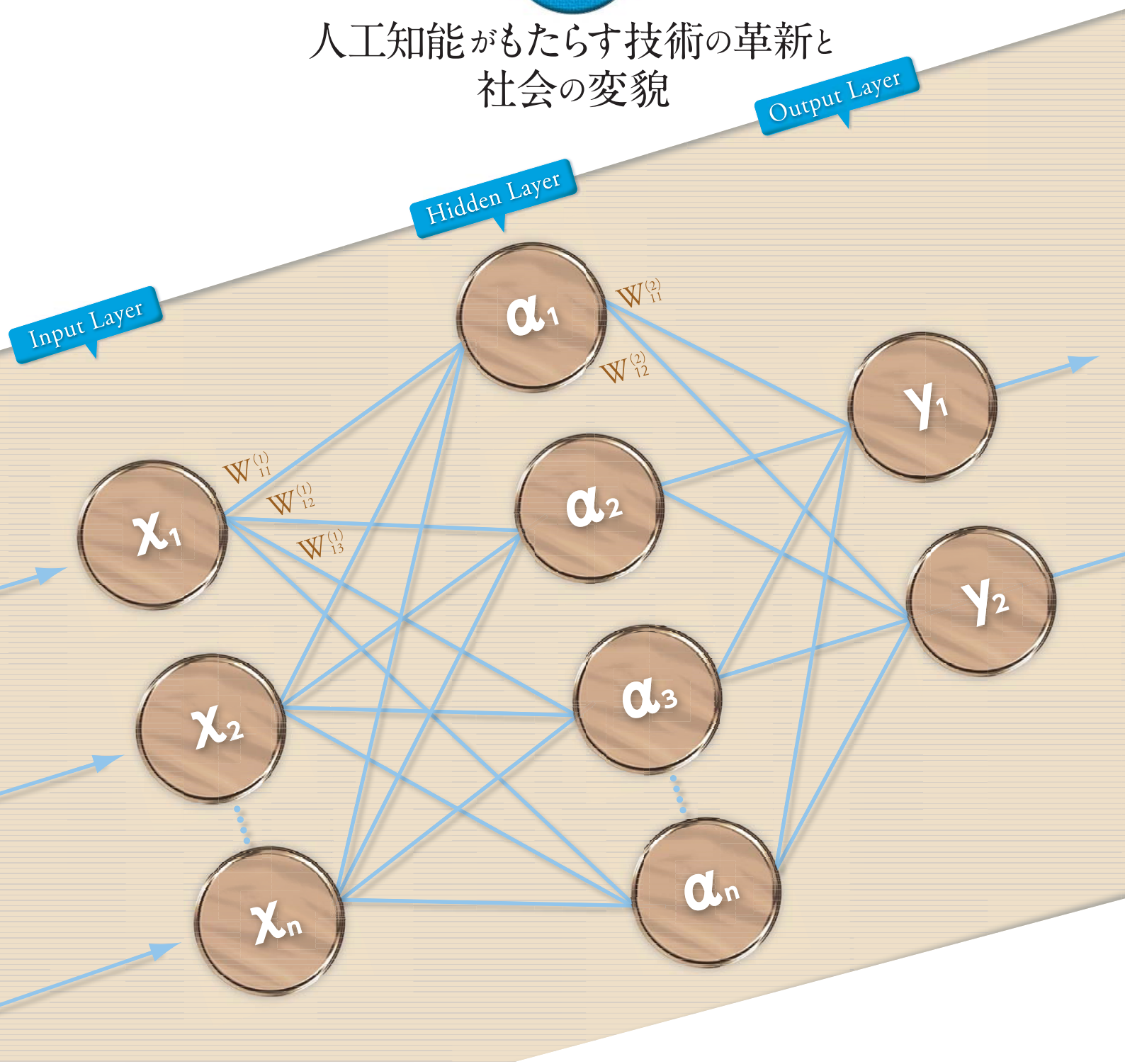
AI白書

Artificial Intelligence White Paper

2017

人工知能がもたらす技術の革新と
社会の変貌



IPA 独立行政法人 情報処理推進機構
Information-technology Promotion Agency, Japan

刊行にあたって

今、世界は人工知能のブームの只中にあるといっても過言ではない程、メディアを毎日のように賑わせています。人工知能によって世界が大きく変わることが期待されています。

人工知能はこれまでも何回かのブームはありましたが、一時的な流行に終わり、社会に大きな変革をもたらすことはありませんでした。しかし、研究としては続けられ、計算機能力の飛躍的な向上、スマートフォンやセンサの普及による大量データの入手の容易性が加わり、さらにはディープラーニングの登場により大きなブレークスルーが起きたと考えられます。

しかし一方で、ブームにありがちな、技術の正しい理解をしないままの様々な誤解が生じています。人工知能があれば何でも簡単に答えが出せると思っている人もいれば、人工知能によって人間の仕事が奪われると拒否反応を起こす人もいます。こうした多くの誤解や思い込みをそのままにしておくと、安易な利用によりうまくいかないケースや、できることも敬遠して利用が進まない状況に陥ります。そうこうするうちに、海外勢による人工知能を利用した新ビジネスが世の中を席巻してしまうことにもなりかねません。まずは、正しい技術の理解と関連情報について、わかりやすく解説した書物が必要ではないかということになりました。そうした背景のもとで、このたび、人工知能に関わる多くの著名な先生方、関係者のご協力を得て、『AI白書 2017』を刊行する運びとなりました。

本書は技術の専門書ではありません。技術そのものに関心のある方には専門書が今、たくさん出ていますので、そちらを探されることをお勧めします。むしろ、正しい技術理解を踏まえて、歴史的な推移を含めた技術動向の今と未来、人工知能の利用動向、人工知能によってどんな素晴らしいことが可能になるのかの実例紹介、人工知能に関わる制度的基盤や国内海外の政策動向といった、人工知能をとりまく全体像を理解して頂くことを目的としています。本書により、一人でも多くの方が人工知能というものについての正しい理解をされ、人工知能を正しく利用される方が増えていくことで、産業が活性化し、日本発の、社会にイノベーションを起こすようなビジネスが多く生まれていくことを期待しています。

最後に、本白書を取りまとめるにあたって、調査や執筆のご協力をいただいた皆様や、編集委員会において、ご尽力いただいた皆様方に対し、心から敬意を表するとともに、厚くお礼申し上げます。

ＩＰＡ（独立行政法人情報処理推進機構）
理事長
富田　達夫

巻頭言

ディープラーニングが起爆剤となった、三度目のAIブーム

2017年はAIの年と言って良いほど、AIブームの真っ只中である。このブームは深層学習（ディープラーニング）の成功が起爆剤になっている。それまでは10年以上先のことと思われていた、コンピュータ囲碁のプロ棋士に対する勝利が報告された2016年から、この動向は急速に顕著になった。これはAIの三度目のブームと言われている。三度目の正直ではないが、今回は定着すると我々は考えている。

前回のAI白書が作られたのは1994年（『AI白書〈1994〉人工知能の技術と利用』通商産業省機械情報産業局電子政策課 監修、日本情報処理開発協会 編）、ちょうど二度目のAIブームの頃である。このブームは、日本で第五世代コンピュータプロジェクトが走っていた1980年代から1990年代とほぼ一致する。当時、人間の熟練者の知識をAIに移植するエキスパートシステムの研究が盛んであった。しかしながら、実用には至っていない。その最大の理由は、人間は言語化できない知識（暗黙知）を持っているからである。その部分が書き下せないため、コンピュータプログラムに移せなかったのである。

現在、深層学習という手法が実用化され、この暗黙知の部分を学習できる可能性が出て来た。ちょうど、人間が弟子入りして先輩たちの熟練の技を見ながら会得するようなことが、コンピュータにできるようになってきた。プログラムとして明示しなくてよいのである。従来型の記号知識と深層学習を組み合わせることにより、AIの発展の可能性が見えてきた。

深層学習の進化は急速である。昨年使った講義資料は、今年は古くてもう使えない。変化は技術だけではない。AIが新しいステージに入ったことで技術革新が加速され、社会の仕組みが急激に変わろうとしている。この時期にAI白書を纏める意義は、大きいと考える。

この白書は時機を逸しないため、約4か月という、通常ではありえないような短期間で仕上げることになった。日本の金融工学の先駆けである今野浩氏による『工学部ヒラノ教授』（新潮文庫）という本に、「工学部の教え7箇条」というのが出ている。「第1条　決められた時間に遅れないこと（納期を守ること）」に始まり、「第7条　拙速を旨とすべきこと」で終わるものであるが、この第7条が特に重要だと考えている。

芸術に限らず、ほとんどの仕事は完成度を追求すればキリがない。本白書もこの方針に則り、拙速を旨とさせていただいた。時機を逃さないことが最重要と考えたからである。言うまでもないことだが「拙」を目指したのではない。「速」を最重視したのである。1年後に白書を編纂すると、今回とはまた異なった風景が見えているに違いない。

AI白書 2017　編集委員会 委員長

中島 秀之

目次

本書のポイント

『AI白書 2017』は、全部で4つの章で構成される。

第1章でAIの最新技術動向を明らかにしたあと、第2章の利用動向ではその技術をもとにどのような利用が、すでになされているのか／考えられているのかを解説。第3章では知財などの制度面での課題とその対応の状況について、そして第4章では国内はもちろん、海外での政策面での取り組みを紹介する。

各章に記載されている内容のポイントを、以下で解説する。

第1章 「技術動向」

ポイント

- 「ディープラーニング」（深層学習）の進展によって、音声・画像認識等のパターン処理では、人間を上回るレベルの認識精度が達成されつつある。
- ディープラーニングによる画像認識は「目」の技術であり、生物が目を得た時と同じく、ロボットや機械の世界でも“カンブリア爆発”的なインパクトになり得る。
- AI及び脳科学等の研究者層の厚みを背景とし、リアル空間のデータを持つ製造業の強みを利用したビジネス開発など、我が国の既存の強みを活かした戦略が求められる。
- ディープラーニングで必要とされる計算インフラの供給によって研究開発・産業応用を加速し、事業開発による利益創出と技術への再投資のサイクルを構築していくことが必要。

概要

ディープラーニングによるAIの進展、ハードウェアの研究開発の活発化

AIの研究開発は、記号的処理の研究からスタートした。言語の発明と使用が、高度な知的社会を人間が構築できた理由であり、記号的処理からAIの研究がスタートしたのは自然な流れだ。だが、こうした方向においてはパターン処理がきわめて弱く、特に視覚的な入力の処理の問題は顕著だった。けれども、ディープラーニングは画像認識や音声認識で大きなブレークスルーを起こした。人間の認識精度とほぼ同等、もしくは超えるという、目覚しい精度向上が実現された。

このディープラーニングによる精度向上を受けて、機械学習用のハードウェアの研究開発も活発化している。特に、ディープラーニングで、ビッグデータに基づいた学習をさせる際には大規模な計算が必要となり、汎用のCPUとは異なる専用の計算用ハードウェアの開発も活発化している。

パターン認識と記号的処理の融合に向けて

ディープラーニングを利用したとしても、現状においてすぐに汎用的なAIが実現するという訳ではな

囲碁AI「AlphaGo」が世界最強のプロ棋士に勝利

Google DeepMindが開発したAI囲碁プログラム「AlphaGo」(アルファ碁)は、2017年5月、世界最強とも言われる中国のプロ棋士、柯潔(か けつ)と対戦。3番勝負で3連勝し、中国囲碁協会から名誉九段を贈呈された。(写真提供:Google)

く、継続的な研究開発が必要だ。今後のAIのソフトウェアに関る研究開発の方向性として、ディープラーニングを基盤的なアルゴリズムとして活用する記号的処理と、暗黙知の処理の統合が挙げられる。

ディープラーニングによる機械翻訳をはじめとして、推論や計画作成等、この方向性に沿った研究開発が萌芽的には存在している。AIの研究開発を支えるハードウェアの研究開発の方向性としては、既存のアーキテクチャのデバイスだけでなく、脳の仕組みを模倣することにより、計算速度向上や消費電力低減を図る試みが挙げられる。これらの研究開発の成果が相乗的に効果を発揮し、AIが継続的に発展していくことが期待されている。

今後、AIのどのような応用がどのような順番で実現していくかについては、必要とされる記号の意味がどの程度深いものであるかによって決まる。画像認識に代表されるパターン認識と、ロボットの動きの学習などは、さほど記号の深い意味に踏み込まずに処理できる領域であり、研究開発と社会実装が進むものと考えられる。その後については、記号の意味を実世界の事物へ関連付ける「シンボルグラウンディング」というハードルを乗り越えることが必要となる。

シンボルグラウンディングは、AI分野で長い間困難と認識されてきた課題であるが、ディープラーニングによる新しいアルゴリズムと、実世界や人間とのインターフェースを持つロボティクスの組み合せにより、解決への糸口が見えて来たところである。

我が国の強みを活かした研究開発が期待される

ネット上のビッグデータに基づくディープラーニングへの取組については、米国等のIT企業が先行した。だが、今後必要となるのは、機械学習用のハードウェアやロボティクス等の、我が国が強みを持つ分野が威力を発揮し得る技術である。現在、経済産業省、総務省、文部科学省が連携して研究開発を推進しているところであり、AIの研究開発が大きく前進することが期待されている。

第2章 「利用動向」

ポイント

●AIの利用には、質の高い学習用データとそれから生成される優れた学習済みモデルが重要。
●学習用データセット、学習済みモデル等を公開・共有し、集合知による加速度的な連鎖が生じている一方、それらを独占する、またAIをデータ獲得の武器として利用する動きも生じている。
●自動運転や医用画像の診断支援等が先進事例。言葉の意味理解に基づく事業創出に向けて、さらなる環境整備（人材、計算資源、標準化等）が必要。
●IoTによって実空間から得られるデータが、AIの今後の競争領域。日本が各産業で保有する強みを活かして、実空間での競争優位を築くことが期待される。

概要

ディープラーニングによるAIの非連続的な革新は、様々な領域で高い成果をもたらしている

ビッグデータの増大とディープラーニングに代表される機械学習の革新により、従来は実現困難だった事業領域で、AIは高い成果をもたらしている。先端的なAI活用は、インターネット空間などの特定の分野における画像認識や音声認識の応用で先行していたが、ディープラーニングの機能を備えたクラウドサービスやオープンソースソフトウェア等の実用性の高いツールの登場により、健康・医療・介護、製造業、金融等の様々な産業や業務領域での適用が進行しつつある。

更に、昨今のビッグデータの増大は、統計的アプローチなど従来から利用されてきた手法の性能を飛躍的に向上させ、適用領域を拡大させている。

先行する企業は集合知のプラットフォームを形成、あるいは学習用データを独占

ディープラーニングに関するアルゴリズム、学習済みモデル、学習用データなど、自らが開発した技術やデータ等を公開し、その上に多様な研究者や技術者を集めて技術開発を加速させる集合知のプラットフォーム形成を進めることで、AIを有効に機能させようとする動きが見られる。一方、質の高い学習用データを獲得、独占することで企業の競争力を高める動きも進められている。将来の市場における優位なポジションを築くために、これらの動きは同時並行的に進むと予想される。

AI活用で米国・中国が先行する中、リアル空間でのAI活用が我が国の競争力向上の鍵となる

AIを牽引する企業は、米国のGoogleやAmazon、中国のBaidu、Tencentなど、インターネット空間を中心に勃興してきた企業である。これらの企業は、スピード感を持って市場にサービスを提供し、ユーザーからのフィードバックを受けながら改善を繰り返し、サービスの品質を向上することで競争優位を築いてきた。今後、AIは様々な実空間での産業に応用され、主戦場は自動運転や医療、介護などの人の命に関わる領域に移りつつある。こうした中、品質・安全性の追求や、ソフトウェアとハードウェ

ア間の機能をすり合わせるノウハウなど、日本が各産業領域で保有する強みを活かして競争優位を築く戦略が求められる。

自動運転を見据えて、すでに多数のカメラ・センサを装備

Teslaがオプションで提供する、レーダーと8台のカメラ、12個の超音波センサ。得られた情報をリアルタイムに処理できる車載コンピュータと、各所に設置されたこれらのカメラやセンサによって、車の全周をカバーする。将来の自動運転を見据えて、段階的にソフトウェアがアップデートされる予定とのこと。（写真提供：Tesla）

第3章 「制度的課題への対応動向」

ポイント

●AIの社会実装の推進にあたって、その存在を想定していなかった既存の法制度等との調和を図る必要がある。
●「知性」という人間の本質に近いところで、「人間の代替」となる側面を持つAIへの不安や懸念に対して、リスクの整理、明確化と、それらへの対応の検討も課題。
●AIが自律的に生成したものは、多くの国の現行法では著作物として認められないが、人間の「創作意図」や「創作的寄与」があれば、著作物性が認められる。
●自動走行システムのガイドラインの整備や、健康、医療・介護分野に関しては、匿名化等の加工をしたうえでの個人データ共有の検討が行われている。

概要

AIの倫理的課題とその制度的対応の議論が、欧米中心に官民挙げて進展

学会やNPOで議論されてきたAIの倫理的課題とその対応について、2016年に入ってから、米国の主要なIT企業（Google、Amazon、Facebook、Apple、Microsoft、IBM等）が中心的な役割を担い、議論を進めている。また、米国、英国は政府や議会でもAIの倫理的課題に関する包括的な検討を実施し、その結果を公表している。更に、IEEEの「ETHICALLY ALIGNED DESIGN」やFLI（Future of Life Institute）の「ASILOMAR AI PRINCIPLES」といった、学界・産業界の幅広いメンバーが参画した団体からAI開発の原則に関する包括的な資料が公表されており、海外では産学官による検討が活発化している。

我が国においても議論は始まっているが、産官学による具体的な検討を加速し、産業応用に即した議論を深めることが求められている。

「AI創作物」や「学習用データセット」「学習済みモデル」の知財面での議論が進行中

AIが創作した音楽や文学作品が出始めている中で、現在の著作権法では保護の対象になっていないこのような「AI創作物」について、「AI創作」の判別可能性や量産性、ビジネス可能性を勘案した保護の在り方について検討が行われている。

機械学習のために、「学習用データ」として他人の著作物等を大量に解析することが著作権侵害か否かが、諸外国において議論されている。我が国の著作権法は平成21年改正によって、コンピュータ等を用いた情報解析のために行われる複製等を許容する、権利制限規定を有している（47条の7）。

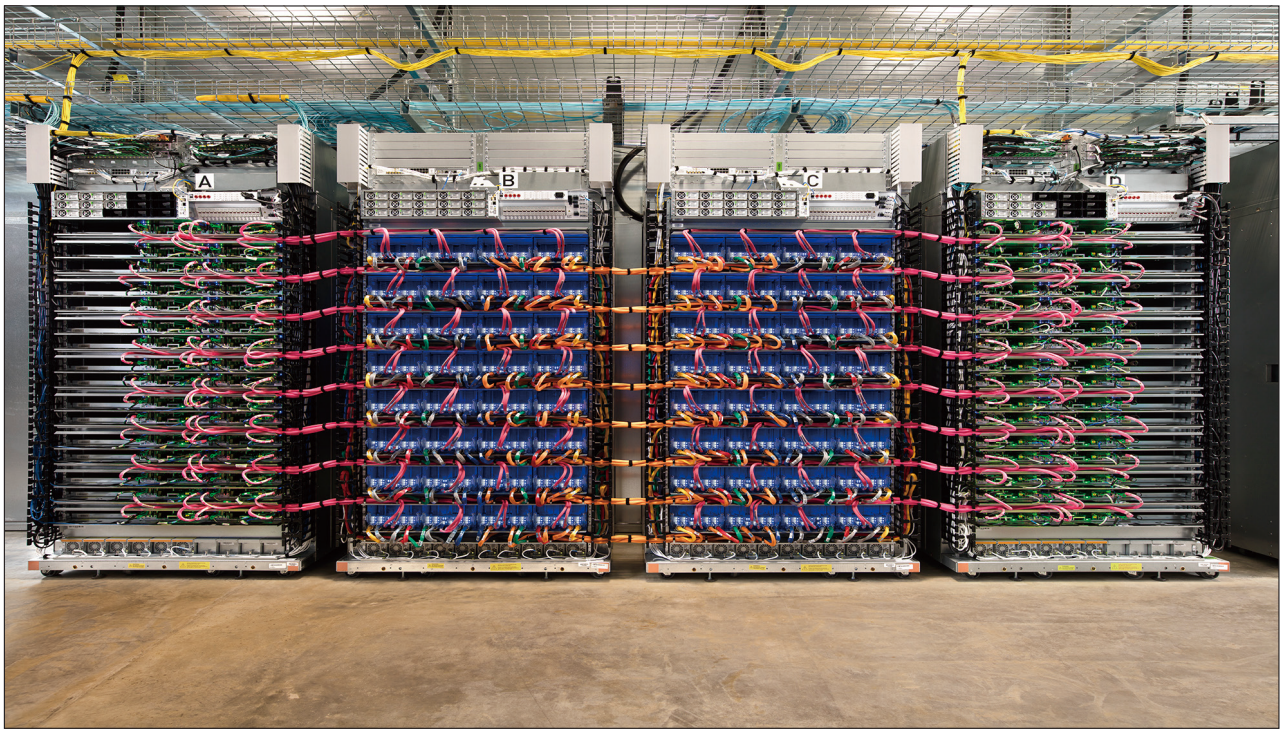
また、「学習済みモデル」の保護の在り方についても、著作権、特許、契約という様々な手段の中で実効性を勘案したビジネス基盤確立に向けた議論が進んでいる。

新たなルール形成が「自動運転」「ドローン」「健康・医療・介護」といった領域で取り組まれている

AIが実現のキーテクノロジーとなっている「自動運転」「ドローン」においては、既存の法体系との整合性を取るべく、法改正や運用ルールの確立に向けた活動が国内外で推進されている。また、健康・医療・介護分野ではAIの機能を十分に発揮できるデータの利活用に向けて、匿名化して活用する仕組みづくりや、パーソナルデータストア（PDS）等の新たな情報活用の仕組みの構築が取り組まれている。

更に、製造業、流通業、サービス業等における物・サービスへのニーズとのマッチングや効率化に向けたAIの活用に関しても、データの匿名化やデータ活用に関するガイドラインづくりが進みつつある。

演算回数ではスパコン「京」を超えるディープラーニング専用機



Googleのディープラーニング専用チップ「Cloud TPU」64個を相互接続した「TPUポッド」。演算能力は11.5PFLOPSで、精度は異なるものの、単純に演算回数の比較では、スーパーコンピュータ「京」を上回る。（写真提供：Google）

第4章 「政策動向」

ポイント

●AIの研究開発に関して、我が国では「人工知能技術戦略会議」が創設され、研究開発目標と産業化のロードマップの策定等が行われている。
●米国では2016年に、AIに関わる研究開発戦略、社会的課題の整理・対応、経済的なインパクトの分析・対応の三つの包括的な報告書が発表された。
●EUでは、欧州全体研究開発プログラムである「Horizon2020」の中で、ドイツでは国策である「Industry 4.0」で、それぞれAIが重要な要素として位置づけられている。
●中国では2016年にAI推進3か年行動計画が策定され、市場創出と研究開発、環境整備がうたわれている。

本白書の記載内容は、原則として2017年4月までの執筆、寄稿、事務局調査に基づく。

第1章 技術動向

第1章 技術動向

1.1 “ディープラーニング”がAIを大きく変えた

1.1.1 AIの研究動向とディープラーニングの登場

人工知能（Artificial Intelligence; AI）分野でのこの数年の大きなブレークスルーは「ディープラーニング」（深層学習）である。海外のAIに関する会議でも、ディープラーニングに関する話題はここ数年で急激に増えた。ディープラーニングがマサチューセッツ工科大学（MIT）の「10 Breakthrough Technologies」（注目すべき10個の革新的技術）に選ばれたのは2013年であったが、画像認識、音声認識、自然言語処理、ゲームなど、様々な領域に飛ぶ鳥を落とす勢いで広がっている。本節では、ディープラーニングに焦点を当てて議論を進めていく。

ディープラーニングとは、深い層を重ねることで学習精度を上げるように工夫した「ニューラルネットワーク[1]」を用いる機械学習（1.2.2項参照）技術のことである。ディープラーニングについては、2006年にカナダ・トロント大学のジェフリー・ヒントン（Geoffrey Hinton）氏らが精度を上げることに成功して以来、様々な手法が提案され、2011年には音声認識のタスクで優勝、2012年にはILSVRC（ImageNet Large Scale Visual Recognition Challenge）という一般物体認識のコンテストで圧勝するなど、数多くのコンペティションで成果を収めてきた。

2013年には、ヒントン氏とその学生らが立ち上げたベンチャー企業DNNresearch（カナダ）をGoogleが買収。そして2014年初頭には、ディープラーニングの先進的技術を有するDeepMind（現Google DeepMind、英国）も、Googleは4億ドル（約445億円）で買収した。

一方、中国のインターネット検索最大手のBaidu（バイドゥ、百度）は、スタンフォード大学でGPU（Graphics Processing Unit）を使ったディープラーニングの研究を進めていたアンドリュー・エン（Andrew Ng）氏を招き、2013年にディープラーニング研究所を設立した[2]。Facebookは、ディープラーニングの主要な研究者であるニューヨーク大学のヤン・ルカン（Yann LeCun）氏をトップに据えてAI研究所を設立し、その後もニューヨーク、シリコンバレー、パリに拠点を広げている。

なお、ICLR（International Conference on Learning Representations）、NIPS（Neural Information Processing Systems）、ICML（International Conference on Machine Learning）などのディープラーニングと関連する国際会議も、ここ数年は急激にその参加者を増やしている。

ディープラーニングがこのように注目されている状況がこのまま続くかどうかに関し、二つの立場からの見解がある。一つ目の立場から見ると、AIあるいはニューラルネットワークに関するブームは、歴

※1
脳の神経回路網で見られる特性を計算機上で再現することを目指した数理モデル。

※2
2017年3月時点でBaiduを辞職している。“Opening a new chapter of my work in AI,” Andrew NG Blog Medium.com Website <https://medium.com/@andrewng/opening-a-new-chapter-of-my-work-in-ai-c6a4d1595d7b>

史的には何回も繰り返されており[1]、現在用いられている手法も、ほとんどが昔からあるものであるため、その用途も現在のところ画像認識や音声認識などに限定されている。したがって、今回が真の突破口であると信ずる理由はないとするものである。

もう一つの立場は、ディープラーニングが今後起こる大きな変化の突破口であると考える立場である。その理由は、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、データを基にして特徴量[3]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からである。

例えば、情報検索の研究は1970年代からあったが、1990年代後半、インターネットの普及という環境を得て、一気に花開いた。インターネット広告という収益の手段を持つ大手の検索エンジン企業でなければ、その後の検索エンジンの研究はもはや事実上不可能になった。産業界における収益化の手段と一旦結び付いた後の学術研究は、（特に米国においては）凄まじい発展を見せる。同じように、ディープラーニングの技術も計算機環境の進展とデータの拡大という環境を得て、一気に花開く可能性が高い。

また、後述するように、このディープラーニングの進展は、日本の強みであるものづくり産業にとって非常に相性が良い。したがって、この技術を活かすことによって、大きく産業競争力を強められる可能性がある。産業界における収益化の手段と学術研究を結び付けることができ、技術が収益につながり、また技術に再投資されるというサイクルを作り出すことができれば、情報技術の領域でここ20年、大きく水をあけられてきた日本も、再び世界に伍することができるようになる可能性がある。

ロボット研究者として有名な、カーネギーメロン大学のハンス・モラベック（Hans Moravec）氏は、著書の中で次のように述べた。

> *The main lesson of thirty-five years of AI research is that the hard problems are easy and the easy problems are hard. The mental abilities of a four-year-old that we take for granted – recognizing a face, lifting a pencil, walking across a room, answering a question – in fact solve some of the hardest engineering problems ever conceived...*（*by Hans Moravec*）

> *35年におよぶ人工知能研究で学んだことは、人間にとって難しい問題は機械にとってはやさしく、逆に人間にとってやさしい問題は機械にとっては難しいということだ。当然のように思う、4歳児の心的な能力――顔を認識したり、鉛筆を持ち上げたり、部屋を横切ったり、質問に答えたり――は、実際、これまでに直面したことのない最も難しい工学的な問題のいくつかを解いている（ハンス・モラベック）*

AIの研究は、大雑把にいうと、1960年代の推論や探索の研究から、1980年代の知識工学・エキスパートシステム[4]、1990年代から2000年代にかけてのオントロジー[5]、セマンティックウェブ（1.5.3項参照）や知識発見、あるいは、マルチエージェント（1.6.1項参照）や社会性、コミュニティといった変遷を経て、2017年現在、機械学習・ディープラーニングの全盛の時代を迎えている。

※3
学習データがもつ特徴を数値化したもの。

※4
ある特定分野の専門知識のデータを基に推論を行い、人間の専門家のような判断を下すシステム。

※5
概念若しくは構成要素の体系化。

これまで、天才ともいえる数多のAI研究者が、知能の問題を考え、そして壁にぶつかってきた。そこで、もっと別のところに答えを見出だそうとする動きが、AIという分野の研究の変遷を生み出してきた。

初期のAIは、記号処理を中心にしたものであり、人間の知能の根源は記号処理にあるだろうというものだった。ところがそれが行き詰まりをみせた。記号処理の中心であったマービン・ミンスキー（Marvin Minsky）氏はMITでAI研究所の初代所長であったが、3代目の所長になったロドニー・ブルックス（Rodney Brooks）氏は「表象なき知能」という概念を提唱し、「服属アーキテクチャ」を考案した。要は、知的に見える振る舞いも、環境との簡単な処理の相互作用の結果として得られるということである。

その後、例えばトム・グルーバー（Tom Gruber）氏は、「オントロジー」という考え方を提唱した。また、知識のもつ社会性に注目した研究が出てきた。複雑性やインタラクションを伴なったマルチエージェントの研究も日本を中心として強い分野となった。

こうした動きは、「群盲象を評す」のように、知能を異なる側面からとらえたものである。ただし、一つだけ足りなかったものが高度なパターン処理であり、それが近年のディープラーニングの進展により解決され始めている。

本節ではまず、AIの分野で、古くから議論されている身体性（1.3節参照）、あるいはシンボルグラウンディング[6]に焦点をあて、ディープラーニングを基盤に置くことでどのように従来の議論をとらえることができるのかを述べる。そのために、「SHRDLU」（次節参照）や服属アーキテクチャ、述語論理などの概念の再解釈を試みる。特に、ディープラーニングにおける生成モデルが、最も重要な要素技術になり得ることを述べる。

1.1.2 ディープラーニングによる生成モデルの可能性

まず、本節での議論を、「SHRDLU」（シュルドゥル）から始める。SHRDLUとは、1968年から1970年にかけてテリー・ウィノグラード（Terry Winograd）氏が開発したシステムであり、AI研究の初期の有名な研究の一つである。

画面の中の「積み木の世界」に、ブロックや円錐、球などが存在し、ユーザからの様々な質問に自然言語文で答えることができた。例えば、「円錐は何に支えられているか？」などである。また、自然言語文の命令により動かすことができた。ユーザは、「緑色の円錐を赤いブロックの上に置け」と指示した後、「その円錐を取り除け」と指示することができた。

だがウィノグラード氏はこの研究の後、AI研究を辞め、HCI（Human Computer Interaction）の研究を行うようになった。次々と先進的な研究を生み出し、研究室からGoogleの創業者まで輩出したにも関わらず、ウィノグラード氏が自然言語処理の研究を辞めてしまったのは、とんでもないほどの絶望感を感じたためである可能性がある。

積み木の世界は全て人間が設計した世界であり、ありとあらゆるお膳立てをして、ようやくコンピュータは少し知的に見える振る舞いをできるようになる。このような裏の事情を分かっている研究者は、一見華やかに見える研究成果の裏にある膨大なお膳立てと、現実の人間の知能の間には、呆然とするほどの距離があることを思い知らされる。

しかし、このSHRDLUに代表される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。

※6
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かという問題。

その鍵となるのが、ディープラーニングにおける「生成モデル」である。通常、機械学習においてクラス分類を解くための手法は、「識別モデル」と生成モデルに分けられる。識別モデルとは、データXが与えられたときにXが属するクラスを同定するモデルであり、X自体がどのように生成されたかについては問わない。これに対して、生成モデルはXが生成される過程までを含めてモデル化する。ディープラーニングで良く使われる「畳み込みニューラルネットワーク」（Convolutional Neural Network; CNN）は識別モデルであるが、生成モデルも近年、数多く提案されている。

有名なものには、「VAE」（Variational Auto-Encoder）や、視点を入れた拡張である「DRAW」（Deep Recurrent Attention Writer）がある。また、「GAN」（Generative Adversarial Network）はイアン・グッドフェロー（Ian Goodfellow）氏らが提案したモデルであり、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精度を上げる。これを拡張した「LAPGAN」（Laplacian Pyramid of Generative Adversarial Network）も有名である。これらを使うと、物理世界での動きを「予想」することができるようになる。例えば、ボールがはずむ動きなどを予想することができる。

また、カテリーナ・フラグキアダキ（Katerina Fragkiadaki）氏らは、ビリヤードの球の動きを、「LSTM」（Long Short-Term Memory）（1.2.3項参照）と「CNN」（1.2.4項参照）で学習させた。これは、ある状況で特定の方向に力を加えると、どのようなことが起こるかを「想像」し、行動の計画を立てることができるというものである。

また、オ・ジュンハク（Junhyuk Oh）氏らは、ATARI（米国）のゲームを題材に、アクションを挟み込んだオートエンコーダ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している。それにより、「DQN」（Deep Q-Network）を使ったゲームのスコアが向上する。

これらが示しているのは、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ということである。従来のように、人間が細部までお膳立てをすることなく、SHRDLUで目指していたような「どういう行動をすれば何が起きるか」をシミュレートすることができ始めているのである。

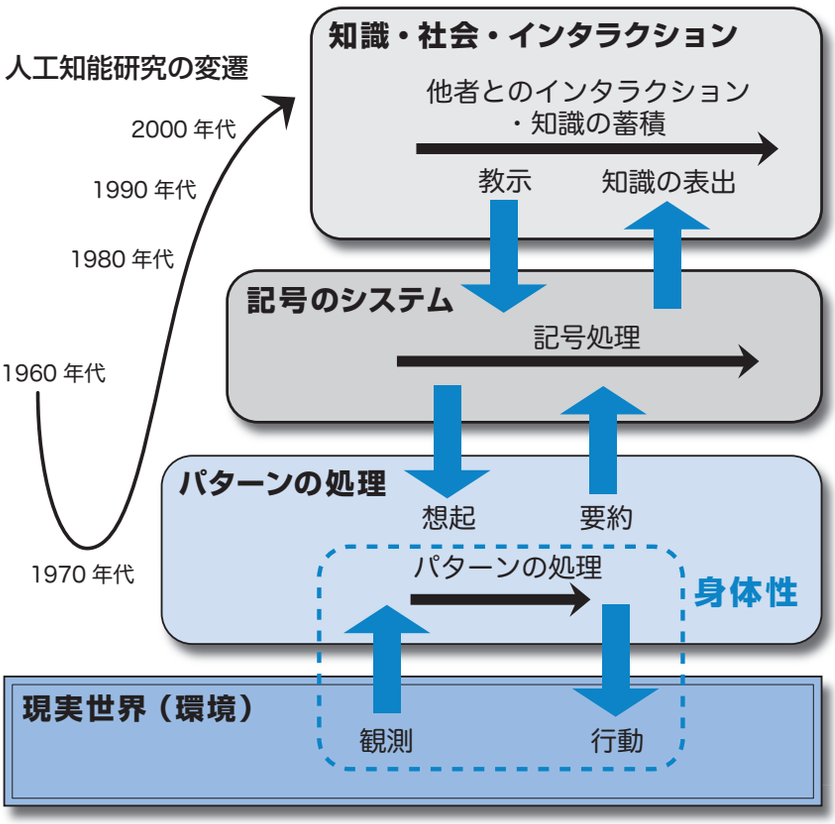
「生成モデルで世界をシミュレートする」というのは単純なアイディアだが、これをベースにして、様々な可能性が広がっている。以降では、身体性への拡張（センサ情報だけではなくアクチュエータ[7]の情報も含んだ拡張）を述べ、その上で、自動翻訳（生成モデルで作った世界と言語との結び付きによる言語の意味理解）、更には、述語論理等による推論（生成モデルで作った世界の記号的な要約）について述べる。

1.1.3 知能の全体像

知能の全体像は、**図1**のようになる。まず、人間も動物も、生物は全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に動くようにも設計できる。これは、ブルックス氏が言っていた身体性であり、ロルフ・ファイファー（Rolf Pfeifer）氏の言う環境における身体性の重要性である。全ての生物は環境に条件付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。

そして、人間の場合は、このシステムの上に記号のシステムを載せている。これはまさに「象はチェスをしない」としてブルックス氏がミンスキー氏を批判したことの裏返しであるが、人間は記号を使っ

※7
入力されたエネルギーを物理的運動に変換する機械要素。



■図1　知能の全体像とAI研究の変遷

たチェスというゲームをする。言語の発明と使用は、外部記憶とその共有を可能にした。このことが、人間以外の種がなし得なかった高度な知的社会を人間が構成できた理由であり、また、多くのAI研究者が当初、自然言語や推論の研究を志向した理由でもある。したがって、ここからAIの研究がスタートしたのは、自然な流れといえる。

なお、記号といっても様々なレベルがあり、内的な記号と外的な記号（コミュニケーションに用いる記号）を分ける場合もある。例えば、我々が次の打合せに移動しようという行動計画を立てる場合には、ミクロなレベルの身体の動かし方から、マクロなプランニングまで、異なる階層で行うわけである。だが、少なくとも上の方はかなりの部分、記号を用いた処理を行っているように思える。なぜ記号が、良い行動選択や学習速度の向上につながるのであろうか。記号はある種の再サンプリングや条件設定の効果、あるいは転移学習[8]の効果があると考えられるが、機械学習の観点から見た記号の意義は、未だに明らかにされていない。

そして、その上に、記号を用いた他者とのコミュニケーションがあり、知識の蓄積がある。この両者はほぼ同じ現象の表裏であり（例えば、オントロジーとコミュニティは同じものであるという議論がある）、知識の取得や蓄積の側に目を向けると、情報検索や知識抽出で研究されてきた分野に当たる。コミュニケーションの側に目を向けると、ソーシャルネットワークやソーシャルメディアの研究ということになる。

ここで述べた全体像は、必ずしも全てのAI研究者が同意するものではないが、大まかにはAI研究者の見方の概観を表すものである。

※8
新規タスクを効率的に処理するためにほかのタスクで学習した知識を適用することで得られる効果。

ところが、こうした知能の全体像において、これまでの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。ブルックス氏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処理の問題は顕著であった（それが今、ディープラーニングにより画像認識の精度が大きく向上し、大きく進展しようとしている）。

したがって、これまでのAIの研究は、記号の研究からスタートし、それが立脚すべき身体性の重要性に思い至り、そこまで遡ったところまではよかったが、そこで解かれるべきパターン処理が技術的な限界によって解けなかった。このため、答えはほかにあるのかもしれないということで、研究の方向は外へ外へと向かって行った。特にマルチエージェントや社会性といった最近のAIにおける研究は、かなり外に進出しているものである。これを**図1**における矢印で表している。

ディープラーニングは、様々な問題の根源的な原因を解決するものであり、そこを起点にして、様々なイノベーションが起こっていくはずである。順番としては、認識の問題が解決されれば、次は身体性のはずである。それが動物としての基本機能だからである。その次に、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」記号を使っての研究である。更に、知識の抽出や共有、コミュニケーションといった話題が出て来て、現象のモデル化の能力も研究されるようになるだろう。

グッドフェロー氏やヨシュア・ベンジオ（Yoshua Bengio）氏らの書いた『Deep Learning』という本[2]には次のような一節がある。

> *One may wonder why deep learning has only recently become recognized as a crucial technology though the first experiments with artificial neural networks were conducted in the 1950s. ... The learning algorithms reaching human performance on complex tasks today are nearly identical to the learning algorithms that struggled to solve toy problems in the 1980s, though the models we train with these algorithms have undergone changes that simplify the training of very deep architectures. The most important new development is that today we can provide these algorithms with the resources they need to succeed.*

> *人工ニューラルネットワークの最初の実験が1950年代に行われたのに、なぜ最近になってようやく、ディープラーニングが極めて重要な技術と認識されるようになったかは、不思議に思うかもしれない。（中略）今日、複雑なタスクで人間の性能に到達する学習アルゴリズムは、1980年代におもちゃの問題を解くのに苦労した学習アルゴリズムとほとんど同一である。これらのアルゴリズムで訓練するモデルは、とても深いアーキテクチャでの訓練を簡略化する変化をしてはいるが。最も重要な新しい進歩は、今日ではアルゴリズムが成功するのに必要とするだけのリソースを、アルゴリズムに提供することができることである。*

つまり、計算リソースやデータが足りなかっただけで、アプローチは間違っていなかったのである。ディープラーニングが画像認識や音声認識で大きなブレークスルーを起こした後にくるものは、ブルックス氏やファイファー氏らが間違っていなければ、身体性の研究しかない。そして、その後にようやくミンスキー氏の世界が来る。したがって、今後は、必然的にロボットの研究が重要だということになる。もちろん、これまでのロボットの研究と異なるのは、ディープラーニングの進展とこれまでの研究が融合されるような形で研究の進展が起こってくるはずであるということである。

1.1.4 ロボット研究の難しさとチャンス

もちろん、ロボットを用いた身体性の研究には難しいところが沢山ある。ディープラーニングのロボットにおける研究としては、今、カリフォルニア大学バークレー校（UCバークレー）のピーター・アビール（Pieter Abbeel）氏らを中心に、ディープラーニングと強化学習を組み合わせるアプローチが研究されている。実ロボットを使い、様々なマニピュレーション、歩行、飛行などのタスクを行っている。もう一方の雄はGoogle DeepMindで、「AlphaGo」への応用が有名であるが、DQNなどの技術が中心に研究されている。3Dの迷路やシミュレータ上でのロボットの研究も行われている。

ほかにも、カーネギーメロン大学やミシガン大学、国内でも早稲田大学や中部大学、プリファードネットワークス（PFN）など、様々な機関で研究が行われているが、世界的には、UCバークレー系と、DeepMind系とにざっくり大別することができるだろう。

両者の戦略の違いは面白い。DeepMindの戦略は、とにかくオンライン空間上でできることをターゲットにするということだろう。CEOのデミス・ハサビス（Demis Hassabis）氏の講演や記事から、脳の処理における身体性の重要性は、明確に理解していると思われるが、それを実現するには、オンラインで行くほうが近道であるという戦略であろう。

一方、Facebook AI研究所のルカン氏が最近、「実世界の限界は実時間でしか動かないことである」と度々言っており、ごく当たり前のことであるが、非常に重要な指摘である。例えば、AlphaGoは1秒程度で1局、自己対局をしているが、オンラインのほうが圧倒的に試行錯誤を重ねることができ、結果的に研究が早く進む。

一方で、UCバークレーのアビール氏らは、実世界を対象に研究を進めている。当然、試行回数を減らさなければならず、よい初期値を与えることが重要で、そのための様々な工夫をしている。行動をいかにチャンク化[9]するか、複数の階層のプランニングを行うか、ある行動を別の機会に転移させるかなどの研究課題がいろいろとある。シミュレータを上手に作る、壊れないように行動を制御する、行動結果を予測するなどの研究も進んでいる。

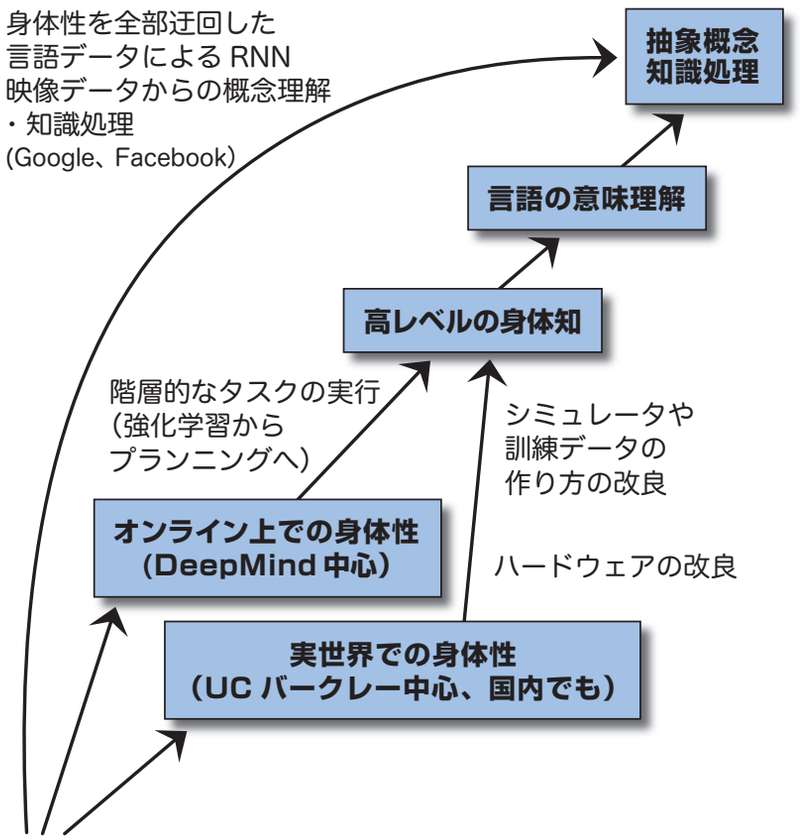
やはり、実世界を対象にした試行錯誤を減らす研究こそが、身体性の獲得においては重要なのだろうか。実世界をいかにシミュレータでモデル化しようが、オンラインのゲームを使おうが、根本的な難しさである実世界の複雑で創造的な非線形性に立ち向かうには、「ほんもの」の実世界を対象にするしかないのではないか。あるいは、「高速に試行錯誤できる」環境で、アルゴリズムを極めた上で、おもむろに実世界を対象にするほうが結局は近道なのではないか。結果的にどちらのアプローチに軍配があがるのか、大変興味深い問題である。

既にその端緒が見て取れるが、こうした研究は次第に記号処理に近づき、その先に、言語との融合があるはずである。言語との融合も、これまでにロボットの研究として様々に行われてきたが、ディープラーニングをベースにした方法に置き換わっていくだろう（ただし、考え方自体は昔から研究されているものと別段大きく変わるわけではない）。

このように、「普通に」考えると、ディープラーニングのブレークスルーから、身体性を介した概念獲得、言語理解というふうに進むはずである。したがって、ロボットの研究をしない限り、AIの研究としては先に進めないはずということになる。

この点においては、日本の研究にもチャンスがある。ディープラーニングの研究を行っているGoogleやFacebookの研究者、あるいはそれを応用しようと目論むシリコンバレーの開発者たちは、実

※9
ひとかたまりのデータとしてカテゴリ化すること。



■図2　身体性をめぐるディープラーニング領域の戦況

世界でのロボット技術に対して大きな苦手意識がある（というより別分野だと思っている）。オンラインで完結する対象（例えば音声認識や画像認識や自然言語処理）に対するディープラーニングの研究者の圧倒的な厚みと、その成果のおそるべきスピードに比べると、ディープラーニングとロボットや機械の交わる領域を対象とする研究者は少なく、日本の研究も十分に戦える余地があると感じる。考えてみれば、ロボット研究は、ハードウェアとソフトウェアの知識・ノウハウが必要とされ、一朝一夕にできるものではないので、ディープラーニングが音声・画像・言語処理を席巻してきたようにはいかないものと考えられる。

こうした研究の進展の順序と、その中での日本の（暫時の間の）ロボット研究に基礎を置くAI研究の優位性が信じられてきた。ところが近年、こういった実世界における身体性という話を全部迂回して、大量のデータから概念獲得や意味理解ができてしまう可能性があるのではないかという危惧も生じ始めている。例えば、体を動かせない人でも、大量の動画を見続けるだけで、現象に対する理解に至るのではないだろうか。世界を理解するのに、本当に「実世界の」身体が必要なのだろうか（この議論は、哲学者フランク・ジャクソン（Frank Jackson）氏が提示した「メアリーの部屋」[10]として知られている）。

GoogleやFacebookは、巨大な言語データも映像のデータも持っているので、そこからの学習だけで相当なところまで到達してしまうのではないか。2016年秋には、Google翻訳がディープラーニングバージョンにアップデートされた。8層のLSTMに、意識的な注意のモデルを入れており、まさに現時点でできる限りの最新の技術を全て取り入れた結果、驚くほど精度があがっている。本来は、実世界の認識、身体性なしには、それほど精度はあがらないと考えられていたが、実際にはかなりあがっている。

この路線が、本当の意味理解に到達する可能性はないに等しいが、GoogleやFacebookの持つデータ

※10
白黒の部屋で生まれ育った女性という設定で行われる哲学的思考実験。

量と、オンラインで完結する対象に対する研究者たちのスピードは凄まじい。こうしたアプローチに、多少なりとも画像や映像などの実世界情報のアラインメント、あるいは獲得した身体性に関する概念とのアラインメントの要素が入ってくれば、大幅にショートカットして、言語理解までたどり着く可能性がある。

その状況を示したのが**図2**である。普通に身体性から攻めると、UCバークレー路線とDeepMind路線の戦いとなるが、大幅にショートカットするGoogle、Facebook路線もあるかもしれないということである。

1.1.5 産業にとっての重要性

こうした状況に対して、日本の研究の立ち位置はどうであろうか。ディープラーニングの研究では日本は相当に遅れている。ICLR、NIPS、ICMLといった主要な会議でも、ディープラーニングに関するテーマで論文を通している日本の研究者はごく僅かである。ただ、身体性という観点から技術を産業と結び付けることにより、その状況を打開することができる可能性はある。

Googleの創業者であるセルゲイ・ブリン（Sergey Brin）氏やラリー・ペイジ（Larry Page）氏はスタンフォード大学の学生であったが、彼らは大学を辞め、Googleを作り、あっという間に時価総額世界一の企業になってしまった。次いで2002年頃には、Facebookが出てきて、あっという間に世界的な企業になってしまった。2017年4月には時価総額で世界5位にまで成長した。

そして、GoogleやFacebookなどの企業は、大量にウェブのエンジニアや研究者、あるいはAIの研究者を雇用している。Googleの年間の研究開発費は１兆円を超える。どう考えても勝ち目がないほど、GoogleやFacebookの研究者の層は厚い。つまり、事業を作り出し、そこから技術に再投資するサイクルを作ったものが結局は勝つということである。それがここ20年で最も激しかったのがウェブの領域だったのではないか。そして、それはディープラーニングの研究についてもあてはまる。

したがって、この状況を打開しようと思えば、ディープラーニングをベースに大きく利益を生み出すような事業を作り出し、技術に再投資するしかないわけである。その観点からは、ディープラーニングが破壊的な変化をもたらす領域がどこか。それを見極めることが極めて重要である。まずは画像認識の領域、例えば、医療画像や監視カメラ、写真の分類やユーザの表情の読み取りなどが対象になるだろう。これらは比較的、シリコンバレーのベンチャーでもやりやすいところであり、既に多くの企業が取り組み始めている。

そして、次に起こるのは、身体性の技術の活用である。認識の技術と、ロボット・機械系の技術の融合で新しい付加価値が生み出されるはずである。典型的には、農業、建築、食品加工の領域が考えられる。これまで自動化、機械化するのが難しく、かつ産業として巨大な分野だからである。

産業用ロボットなどの活用もあるが、これまでの工業化はいわば「目の見えない機械」を使って環境を整えることによって自動の処理を実現してきた歴史であった。したがって、今後起こる変化は、環境が整っていないところでの自動化が起こるということである。そのなかで産業として大きいのが、農業や建築、食品加工などであろう。それ以外にも、例えば、片付けというのは巨大な市場を構成するのではないか。認識し、ものをあった場所にしまうというだけだが、それだけで家事労働が非常に楽になるし、生活の質が大きく向上する。

地球ができて46億年だが、5億4200万年前から5億3000万年前という短時間に、現存する全ての生物の門が出揃ったという現象を、「カンブリア爆発」と呼ぶ。これは「眼の誕生」が原因であるとする説を、アンドリュー・パーカー（Andrew Parker）氏が提唱した。そして、ディープラーニングは「目」の技術であり（視覚野の処理を実現するものであり）、ロボットや機械の世界でこれからカンブリア爆発が

起こるということである。TRI（Toyota Research Institute、米国）のギル・プラット（Gill Pratt）氏やソフトバンクの孫正義氏も同じ表現をしている。

こうした片付けや調理、あるいは農業や建設などの作業をロボットで実現することは、ロボットの研究コミュニティでは古くから試みられている。しかし、今がそのタイミングなのではないか。ウェブの世界でも、情報検索という技術は昔からあったにもかかわらず、ウェブページを対象とする検索エンジンで一気に商用化した。ソーシャルネットワークの研究は社会学の分野で何十年も行われていたが、SNSは一気に広がった。つまり、昔からあるアイディアが、何らかの現実的な条件が満たされて一気に広がることはよくある。

技術的には、ディープラーニングにより画像認識の精度が非常に上がり、強化学習との連携が使えるようになってきたことが極めて大きな変化である。あとは、ハードウェアの性能の問題と、製造コストがどこまで下がるかである。ここがクリアされれば、大きな事業につながるかもしれない。そして、ハードウェアの問題、コストの問題は、日本の研究者や日本企業が強みを発揮できる部分である。一旦事業として成立し始めれば、更に精度の良い学習を行うことに対して、事業者は強いインセンティブを持つはずである。それが研究の後押しとなり、技術を大きく前に推し進めることになる。

ディープラーニングを起点にした事業化をうまく進めることができれば、インターネットの分野におけるGoogleやFacebookのような地位を日本が取れる可能性があるのが、ロボット・機械の分野である。したがって、AIの研究者という立場からみて、ロボット研究に対する期待は極めて大きい。

1.1.6 ディープラーニングに基づく記号の意味理解に向けて

環境世界を知覚する仕組みの上に、世界を予測する生成モデルが築かれ、その上に更に記号の操作が実現されるという知能の全体像を知るという課題に対して、現在のディープラーニングの研究の中で、その解明の端緒に当たるものは二つある。

一つは、画像から文を生成する自動キャプション付けであり、画像中で最も注目すべき点に焦点を当てて、文を生成する手法である。例えば、画像を与えると、「ピンクの服を着た女の子が芝生の上でジャンプしている」などの文を生成する。もう一つは、画像の生成であり、文が入力されると画像を生成するようなモデルを学習するものである。例えば、DRAWを使って、「飛行機が空を飛んでいる」「象が砂漠を歩いている」などの文を入れると、該当する画像を描くことができる。「止まれの標識が空を飛んでいる」などの、普通ならあり得ない文を入れても画像を生成できるところが興味深い。これは、すなわち、自然言語文から画像を生成することができ、更には画像から自然言語文を生成することができるということである。再度、SHRDLUの自然言語による操作と同じことが実現できるわけである。自然言語文の「意味を理解」していると言えるかもしれない。

さて、AIの分野では、意味理解についての議論も古くからある。アラン・チューリング（Alan Turing）氏は、チューリングテスト[11]を提案し、「計算機は考えることができるか」という問いを、「模倣ゲームをうまく行うことのできるような想像上の計算機は存在するか」という問いに置き換えた。それに対して、ジョン・サール（John Searle）氏が1980年に示した思考実験が「中国語の部屋」である。仮にチューリングテストに合格する機械ができたとしても、操作している対象の「意味が分かっていない」というものである。

文の意味が分かるとはどういうことだろうか。文の意味を理解するとは、文から画像を生成すること

※11
ある機械がAIであるかどうかを判定するためのテスト。

ができることである。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言い方であり、実際には、センサとアクチュエータの複合的な時系列情報であるので、体験という言葉のほうが適切であろう。つまり、意味理解ができるというのは、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のことであると考えられる。

ディープラーニングの生成モデルを使えば、本当の意味での意味理解を行うことができるはずである。例えば、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ができるかもしれない。すなわち、自動翻訳、しかも意味理解を伴う自動翻訳が可能になるかもしれない。もちろん、本格的な自動翻訳を実現するには、沢山の課題がある。

- 抽象的な概念をどのように扱うのか。人間の意味理解が、視覚情報や視覚的な処理機構をベースにしているのは確かにそうであるとしても（抽象的概念でも空間的な扱いをするものが多い)、映像として再現することは難しい概念も沢山ある。センサ、アクチュエータの高次の特徴量が復元されるということでいいのだろうか。
- 感情や本能等に関わるものをどのように扱うのか。例えば、美しい、おいしいといった感覚は学習できるにしても、人間と同じような感情が実現されているわけではない。その設計をしなくとも（例えば納豆が嫌いでほとんど食べたことがない人でも、納豆を食べる人を観察して学習できるように）ある程度何とかなるのだろうか。
- 人間と同じセンサ、アクチュエータがないと、人間と近い（あるいは理解し得る）概念を生成することは難しいのか

もう一つ、考えなければならないのは、記号処理により、見えていないことをいかに予測するかである。AIで長らく研究されてきた命題論理や述語論理、あるいは様相論理などによる推論[3]は、与えられた知識や事実から、最初は見えていない帰結を導き出すための仕組みであった。その点では、最近、注目を集めたAlphaGoの研究も意義深い。過去の棋譜データや自己対戦データを用いながら、CNNを用いた上で、「policy network[12]」、「value network[13]」を構成していく。それによって、先読みを大幅に効率化している。

そもそも、こういった思考ゲームがコンピュータに扱いやすいのは、世界モデルを構築することなく、シンプルなルールを記述しておくだけで、未来の状態を展開することができたところにある（つまり生成モデルによる世界のシミュレートを「さぼる」ことができたわけである)。ところが囲碁においては、一手一手の操作があまりにもプリミティブすぎて、結局は世界モデルの構築が重要な鍵であった。それに対して、AlphaGoでは、CNNによる盤面の認識に加えて、強化学習によるpolicy networkを構成することで、探索する範囲をかなり絞ることに成功した。

おそらく人間の場合は、視覚的な生成モデルをベースにしながら、こういうときにはこうなるという関係性を記号レベルの連接関係でも学習していく。すると、いちいち重い処理が走らずとも、簡略化して思考を先に走らせることができる。この記号の想起と、視覚的な生成モデルの組合せが、思考の過程であり、それを（生成モデルによる世界のシミュレータがないがゆえに）シンボルの想起だけに限定したものが、従来の述語論理や様相論理による推論ということができるのではないだろうか。

※12
次の手を選択・評価するための畳み込みニューラルネットワーク(Convolutional Neural Network; CNN)。

※13
局面を評価するためのCNN。

1.1.7 本章の構成

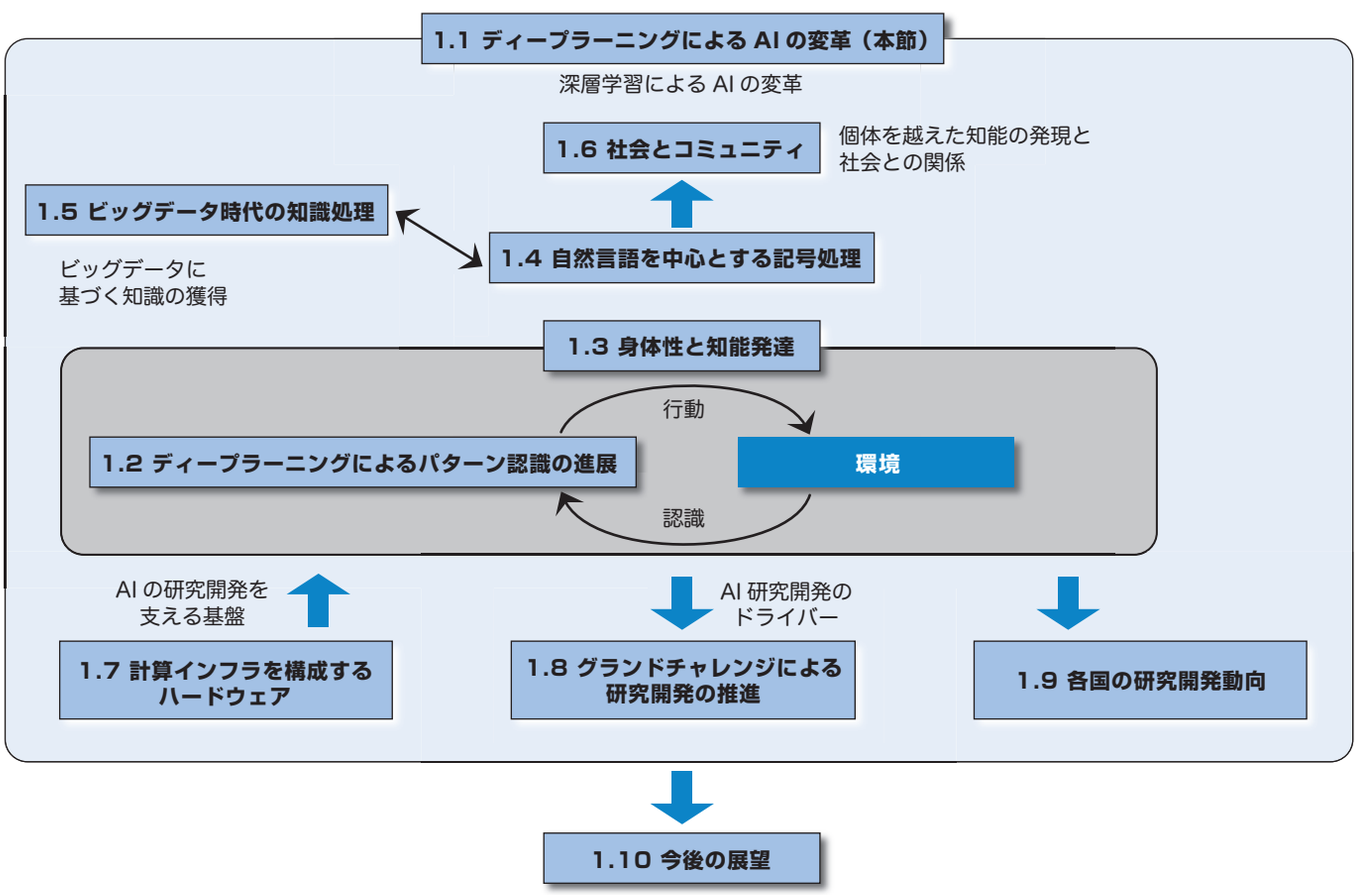
上記の状況認識を踏まえ、本章は次のように構成する。

まず、ディープラーニングを中心とする無意識の処理やパターン認識について、**1.2節**で述べる。画像認識で多く用いられているCNNや、時系列の処理を行う「リカレントニューラルネットワーク」(Recurrent neural network; RNN）を中心にその理論や応用について述べる。

次に**1.3節**では、身体性についての研究の動きをまとめる。ディープラーニングの急速な発展をベースに、今後大きく変わっていくであろう技術である。特に、産業的にも日本が強みを発揮する上でこの観点は極めて重要である。

1.4節では、記号・言語についての研究を概観する。記号処理は、まだディープラーニングの影響が大きく及んでいる領域ではないが、機械翻訳を始めとして先進的な動きが進みつつあり、今後、従来からのAI研究と新しい潮流がぶつかる、そして人間の知能を構成する上で大変興味深い部分に研究が突入すると思われる。

1.5節では、こうした進展を支えるビッグデータの技術についてまとめる。ディープラーニングはビッグデータを必要とするが、非常に自由度の高い（capacityの大きな）モデルを学習しており、capacityに比べるとデータ量が少なくてすむような工夫が随所になされている。知能の研究は、できるだけ少ない量で同じモデルを学習する、あるいは、同じデータ量でよりcapacityの大きなモデルを学習するという方向に進んでおり、データの大きさだけで語るのはフェアではない。しかしながら、技術的にはビッグデータの取扱いが大きなイノベーションを起こしてきたのは事実であり、大変重要な技術である。また、ディープラーニングに限らず、ビッグデータによる様々な分析や活用が可能になっており、その流れも今日のAIを特徴付ける一つであろう。



■図3　1章の全体構成

1.6節では、社会とコミュニティについて概観する。AI研究では比較的最近進展してきたトピックであり、社会を構成することが人間の知能を大きく高めたことからも、その重要性は明らかである。ディープラーニングとの融合はまだ先かもしれないが、AI研究においては重要なトピックの一つである。

1.7節では、AIのインフラストラクチャやハードウェアについて述べる。ディープラーニングに用いられるGPUを始めとして、様々なインフラストラクチャやハードウェアがこうしたAIの進化を支えている。省エネの技術や組込みの技術なども、今後、AIが様々な産業で使われる上では重要であろう。

1.8節では、AIのグランドチャレンジを取り上げる。AIの分野では、歴史的にグランドチャレンジが大きな役割を果たしてきた。「ロボカップ」、「DARPA（国防高等研究計画局）Grand Challenge」、「DARPA Robotics Challenge」等である。大きな夢を共有することで、それに向かってコミュニティの力を結集しようというものであり、それによって様々な技術がスピンアウトして出てくる。

1.9節では、我が国及び諸外国の政府と民間企業による研究開発の動向を概観する。

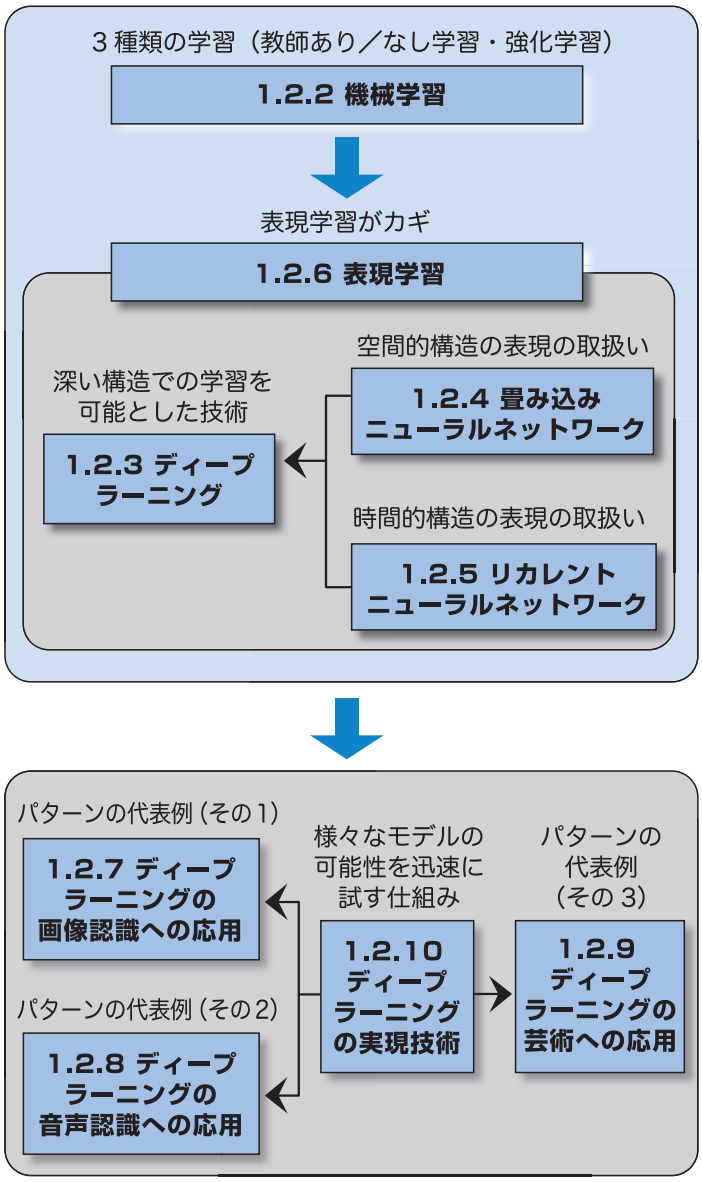
1.10節では、AIの今後の展望を述べる。こういった技術の進展はどのようなステップで進むのか、この先に技術はどこまでいくのかなどを述べる。

参考文献

[1] 麻生英樹ほか『ディープラーニング』近代科学社.
[2] I. Goodfellow et al., “Deep Learning,” *The MIT Press.*
[3] 石塚満『知識の表現と高速推論』丸善.

1.2 ディープラーニングによるパターン認識の進展

1.2.1 総論



■図4　本節の構成

1.1節では、ディープラーニングがどのように人工知能（AI）の枠組みを転換し得るのかについて概観した。本節では、そのディープラーニングについての基本的な考え方と応用に関する現状についてより詳しく述べる。

ディープラーニングは、情報処理の単位であるニューロンが層状に接続した構造を模擬した機械学習の一種である。以下では、最初に、機械学習の三つの枠組みについて述べる（1.2.2項参照）。更に、三つの枠組みの中で、特に身体性（1.3節参照）との関係で重要となる強化学習についての課題について詳述する。

ディープラーニングは、ニューロンが多層で接続しているモデル構造（深い構造）を持つことにその名前は由来している（1.2.3項参照）。脳は、ニューロンがネットワーク状に接続しているため、その構造を素直に数理モデルで定式化したものであると言える。ネットワークを多層構造にすれば、基本的な概念から、その組合せで表現される高次の概念まで表現可能なため、アイディアとしては古くから存在

しており、ファジィシステム研究所の福島邦彦氏による「ネオコグニトロン」は、その先駆的な研究とされている。今回のディープラーニングのブレークスルーは、トロント大学のヒントン氏らによる研究を端緒として、音声認識、一般物体認識を始めとする様々なコンテストで優勝したことによる。一口にディープラーニングと言っても、現在提案されているネットワーク構造は多岐にわたっているが、最も典型的なフィードフォワード[1]の多層ネットワークを例に、その構造と機能を説明する。

現在、データに潜む空間的構造をモデル化する場合と、時間的構造をモデル化する場合で、よく使われるネットワーク構造が分かれており、空間的構造では「畳み込みニューラルネットワーク」(Convolutional Neural Network; CNN)、時間的構造では「リカレントニューラルネットワーク」(Recurrent Neural Network; RNN）が多く使われている。

CNNは、前述の福島氏によるネオコグニトロンを起源とし、脳の視覚野の構造とも類似の機能を持つネットワークであり、現在のディープラーニングによる画像認識に関して、デファクトスタンダードとなっている構造である（1.2.4項参照)。画像中の物体に僅かなずれやゆがみがあっても吸収する仕組みが導入されているため、頑健な認識が可能となっている。

一方、時間的構造の場合には、RNNが利用される場合が多い。ここで言う時間的構造とは、単純な時系列データの構造だけではなく、論理展開や自然言語、音楽など、順序的な構造を含むデータであれば対象となり得る（1.2.5項参照)。

また、AI研究における「表現」(representation）とは、概念に対応したニューラルネットワークの状態のことを指す。ディープラーニングの登場以前には、表現の材料とも言える「特徴量」を、研究者自身が工夫してあらかじめ作成する必要があった。これに対して、ニューラルネットワーク上での「特徴量」の獲得を可能にしたのが、ディープラーニングである。これにより、モデルを構築する際の拠って立つ基盤となる土台が手に入ったことを意味している。

特に、自然言語に代表される「記号」に対応する表現がどのようなものであり、どのように獲得されるのかという問題は「シンボルグラウンディング問題」と呼ばれ、AIの歴史の中で、解決が困難とされてきた課題の一つである。ここでの「記号」とは、静的なイメージ、動的なイメージだけでなく、情報処理の中間的な状態、動機や感情を伴う状態など、脳内の活動のあらゆる側面を含む。シンボルグラウンディングにディープラーニングが重要な役割を果たすのは確かであると考えられるが、必ずしもその構造が必然的なわけでもなく、今後様々な表現学習の手法が出現する可能性が考えられる。

ディープラーニングの応用分野の中で、画像認識の分野は、パターン認識の結果が視覚的に分かりやすく、画像中の物体検出や物体セグメンテーションなど、自動運転、画像診断、防犯画像の認識等の応用にも直結するため、盛んに研究されている分野である（1.2.7項参照)。生成モデルの研究も画像分野が一番進んでおり、キャプションからの画像の生成等が行われている。また、画像的なイメージを伴う概念は多く、シンボルグラウンディングを実現するためにも重要な分野であり続けると考えられる。

音声認識も、画像認識と並んで実用に直結する分野である。スマートフォンやコールセンターでの利用や、今後AIの活用領域の拡大が進むにつれて、機械と人間のインターフェース（マンマシンインターフェース）に音声認識を組み込むケースが増大すると考えられる（1.2.8項参照)。音声分野における生成モデルとしては、テキストから合成音声を生成する研究も既に発表されており、従来に比べて聞き取りやすい音声が得られたとされている。

また、生成モデルを用いた応用として、芸術等の分野でも研究例が増えている（1.2.9項参照)。画像

※1
フィードバックの逆。制御を乱す外的要因(外乱)が発生して、それが影響として現れる前に、前もってその影響を極力なくすように必要な修正動作を行う。

の生成、音楽の生成をはじめとして、長期的には小説の生成も目指されている。

本節の最後に、ディープラーニングの開発の際に必要となるソフトウェアについて動向を紹介する（1.2.10項参照）。ディープラーニングを実装するためのソフトウェアには、IBMの「Watson」（ワトソン）を始めとする商用のソフトウェアも存在するが、世界のトップクラスの研究開発で使用されているソフトウェアのソースコードの多くは公開されている。その中で、Googleの「TensorFlow」（テンソルフロー）、Berkeley Vison And Learning Centerによる「Caffe」（カフェ）、そして日本のプロフォードネットワーク（Preferred Networks; PFN）が開発した「Chainer」（チェイナー）等が頻繁に利用されているものである。

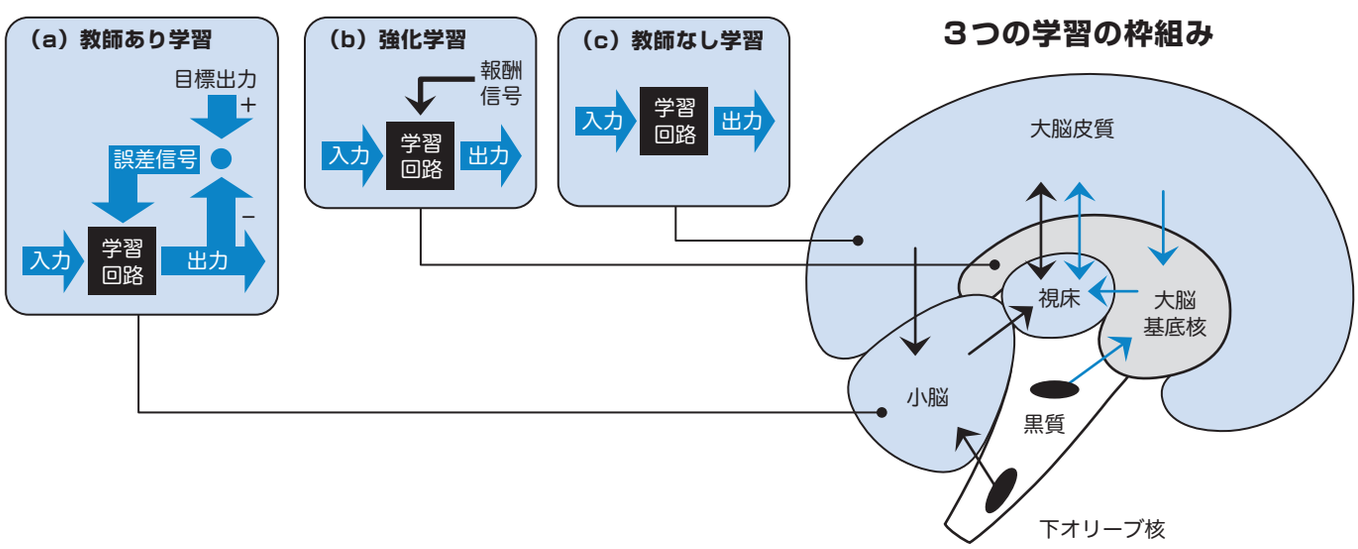
これらのオープンソースソフトウェアには、それぞれに特徴があり、どれか一つでオールマイティというわけではない。GoogleのTensorFlowは大規模なGPU（Graphics Processing Unit）環境での高速化に特徴があると言われている[2]。PFNのChainerは、実行時にも動的にネットワーク構造を変えられる「Define-by-Run」という方式を取っており、ディープラーニングの世界で次々に提案される新しいネットワーク構造やモデルを簡単に取り込むことが可能とあって、人気を集めている。

これらの頻繁に利用されるオープンソースソフトウェアをベースとして、個々の研究者が自身で考案したモデルを実装、公開することで、他の研究者もすぐにそのモデルを試せるといった好循環が生じている。

1.2.2 機械学習

1.2.2.1 機械学習の構造

脳における学習の枠組み[1]に基づき、機械学習における三つの学習の枠組みを紹介する[2]。



■図5　三つの学習の枠組み

三つの学習とは、「教師あり学習」、「教師なし学習」、そして「強化学習」である（**図5**）。これらは、脳の部位として、それぞれ小脳、大脳皮質、そして大脳基底核と深く関連がある。

※2
Martin Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems."
<http://download.tensorflow.org/paper/whitepaper2015.pdf>

- **教師あり学習（Supervised Learning）：**
  主に、小脳が担う。学習者に対し、教示者が明示的に正例を呈示したり、学習者の誤りを指摘したりすることで、学習者が正しい解を得ることを助ける。すなわち、正しい入出力の組合せを与えて学習することで、新規の入力に対し、適切な出力を提示できる。Back Propagation（誤差逆伝播法）[3]が、その代表的手法であろう。また、正例、若しくは負例を入力として、未経験入力に対する意志を決定する決定木（Decision Tree）の作成などもある。
- **教師なし学習（Unsupervised Learning）：**
  主に、大脳皮質が担う。統計的性質や、ある種の拘束条件により入力パターンを分類したり、抽象化したりする学習。主成分分析、自己組織化マップなどの次元圧縮（Dimensionality Compression）手法が代表例である。感覚情報などの入力パターンの分類、同様に出力運動パターンに対して統計的性質を用いて要素行動に分類する学習法などがある。
- **強化学習（Reinforcement Learning）：**
  主に、大脳基底核が担う。最終結果若しくは、途中経過に対して、どの程度良かったかを示す「報酬信号」に基づき、これらの報酬をなるべく大きくするように探索する。

強化学習と教師あり学習の違いは、フィードバックがスカラー（成否）かベクトル（howの情報）かという説明もあるように、明示的な教師ではなく、環境などの非明示的な教師だという解釈もある。

1.2.2.2 強化学習の課題と最近の動向

ここでは強化学習の基本課題として、遅延報酬による長期学習時間（学習時間）、状態行動空間構成（状態行動空間）、マルチエージェント応用（スケールアップ）などについて概説し、最近の動向についても触れる。

（1）学習時間

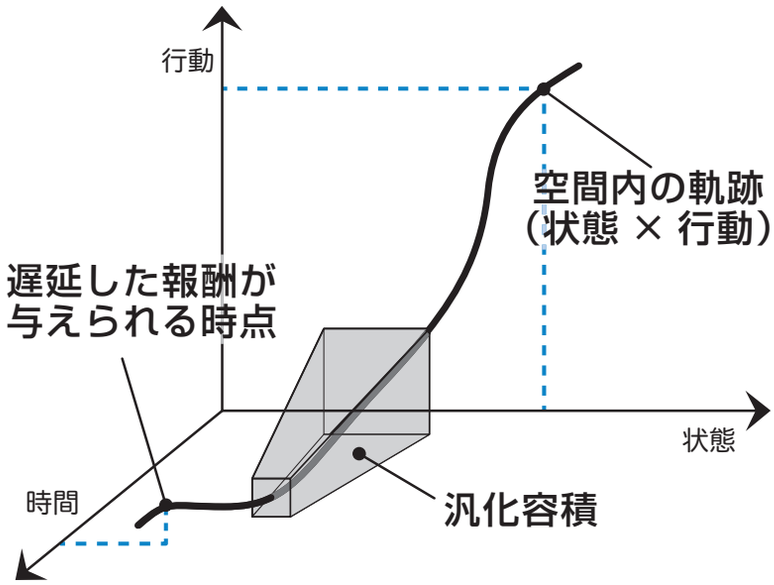
理論的には無限に学習するが、実世界では全てが限られている。ロボットの場合、無限の試行を繰り返すことなどできず、ロボットが損耗し、実験の続行が困難になる。人間の場合でも、何度も失敗が続けば、それこそ動機を失う。そこで、やさしいタスクからの学習（Learning from Easy Missions; LEM）を設定することで、理論的に、探索時間を状態行動空間のサイズの指数オーダから線形オーダに圧縮可能である。先験的にタスクの「やさしさ」が分かれば問題はないが、そうでない場合、その確信のなさに従い、線形オーダから遅くなるが、元の学習の収束性が保証されていれば、同様に保証される。

（2）状態行動空間

状態が格子状で行動が格子間の移動などの理想的な状態行動空間は、イベントベースの抽象的な状態行動空間を除き、実世界ではほとんどありえず、セグメンテーション課題と呼ばれる大基本問題の一つである。報酬が与えられる時間も含めて「クレジット割り当て問題」（Credit Assignment Problem）と呼ばれている（**図6**）。クレジット割り当て問題とは、状態・行動の空間内での軌跡が与えられたとき、遅延した報酬が与えられたとき、過去のどの時点のどの範囲の行動を強化すれば良いのかという問題である。状態行動空間を再帰的に定義することで、状態行動空間構成の「鶏と卵」問題を解消した手法が

※3
正解と実際の出力を比較することで各層間の重み付け等を修正する学習方法。

提案されている。また、初期を一状態とし、連続の状態行動空間を線形関数近似により分割する手法[3]では、線形関数近似に加え、報酬（ゴール到達）の成否による細分化も含まれている。最近では、ベイズ推定の枠組みで、状態・行動空間を自律的に分割する機構を持つ強化学習法が提案されている[4]。



■図6　クレジット割り当て問題[4] [5]

(3) スケールアップ

より複雑なタスクへの応用として、階層構造化とマルチエージェント化の課題が挙げられる。前者では「MOSAIC」[6]が有名だが、高橋泰岳氏ら[7]は、均一な強化学習器を多く準備し、階層のレベルを、それらの能力と環境に依存して（事前に指定しない）、自律的に構造化する手法を提案している。マルチエージェント学習では、同時学習による学習過程の不安定化[5]が課題である。

文献[8]は、最近の発展系などに詳しい。ベイズ統計との関連、報酬関数を未知として、人間のエキスパートの行動から報酬関数を推定する逆強化学習、群強化学習、また、強化学習の関数近似にディープラーニングの技術を適用したDQN（Deep Q-Network）などは、1.1節でも紹介されているように、ゲームや実際のロボット応用にも適用され始めている。内発的動機付けとの関連[9]もロボットの自律性などの観点から興味深い。

1.2.3 ディープラーニング

ディープラーニング（深層学習）は、狭い意味では、層の数が多い（深い）ニューラルネットワークを用いた機械学習のことである。より広い意味では、低次の局所的特徴から高次の抽象度の高い大域的特徴に至る、深い階層構造を持つ特徴表現（内部表現（internal representation）とも呼ばれる）をデータから獲得する機械学習を指す[10]。

複数の中間層を持つ階層的ニューラルネットワークの、結合の重みをデータから学習させるための方法として、1980年代に誤差逆伝播学習が提案され、様々な問題に適用されて一定の成功を収めた。だが、層の数が多いネットワークをうまく学習させることは難しいとされていた。その理由としては、

(1) 出力層における誤差を入力層に向けて伝播させる間に、誤差情報が徐々に拡散し、入力層に近い層では勾配の値が小さくなって学習がうまく進まないこと（勾配消失現象）。
(2) 層の数が多いニューラルネットワークの学習の目的関数は、非常に多くの局所的な極小値（ローカルミニマム）を持ち、適切な結合の重みの初期値の設定が難しいこと。

※4
文献[5]より作成。

※5
初心者二人のテニスを連想するとよい。どちらも下手なので、練習もできない。相手がコーチだと定まったボールを初心者に呈示するので、初心者は安心して練習できる。

などが挙げられる。原理的には、中間層が一つのニューラルネットワークによって任意の連続関数が近似可能であるため、層の数が多いニューラルネットワークを学習させる試みはあまり行われなかった。

2006年頃にヒントン氏らのグループは、「深層信念ネットワーク」や、「制限ボルツマンマシン」[6]を多数積み重ねたオートエンコーダ[7]等の手法によって、様々な種類のデータに対して、深い階層を持つ有効な特徴表現が得られることを示した。更に、2011年頃から、不特定話者の連続音声認識や静止画像中の一般物体認識などの難しいパターン認識タスクにおいて、ディープラーニングによって得られる特徴表現を用いて、従来法を大きく上回る性能が得られたことから、ディープラーニングの手法と応用の両方に関する研究が非常に盛んに行われるようになっている。現在では100層を超える多層のニューラルネットワークも学習可能になっており、特に、一般物体認識や顔の識別などの画像認識系のタスクで、人間と同等以上の認識精度を達成しているものも多い。

画像の認識では、主に入力から出力に向かう結合のみを持つ階層的なニューラルネットワーク、特に、画像などの信号に内在する局所的な特徴が集まって、より大域的な特徴を構成するという構造を反映した、畳み込みネットワークがよく用いられている。一方、自然言語テキストや動画に代表される、構造を持った系列情報を扱うために、RNNも再び研究されるようになった。なかでも、リンツ・ヨハネス・ケプラー大学（オーストリア）のゼップ・ホフレイター（Sepp Hochreiter）氏らの提案した「LSTM」（Long Short-Term Memory）は、必要な文脈情報の長さを適応的に制御することで、時間を遡る誤差逆伝播学習の可能性を向上させる点が再評価された。画像からの説明文の生成や機械翻訳など、多くの課題に適用されている。

パターン認識のための識別モデルとしてのみならず、画像などの観測情報を生成するための生成モデルとして、層の数が多いニューラルネットワークを用いることも研究されている。更に、識別モデルと生成モデルを組み合わせて、相互かつ敵対的に学習させることで、全体の性能を向上させる手法も考案された。

強化学習との組合せや、スタックなどの外部記憶構造との組合せなどによって、ディープラーニングを拡張していく研究も進められている。強化学習とディープラーニングの組合せは、強化学習の性能を大きく左右する状態空間の情報表現を、ディープラーニングによって獲得させられる利点がある。トロント大学のヴォロディームィル・ミン（Volodymyr Mnih）氏らは、古典的テレビゲームに適用して、多くのゲームで人間を超える性能を実現した。カリフォルニア大学バークレー校のセルゲイ・レヴィン（Sergey Levine）氏らは、接触の多い“はめあい”などの動作をロボットに学習させた。ロボットによる物体のピッキングなどへの応用も進められている。

そして、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囲碁ソフトウェアである「AlphaGo」が、世界トップレベルの棋士に勝利するなど、目覚しい成果を挙げており（1.8.1項参照）、今後、ロボット制御やシステム最適化などへの適用が更に広がることが期待されている。また、トロント大学のアレックス・グレイヴス（Alex Graves）氏らは、ディープラーニングと外部記憶構造を組み合わせたモデルの全体を学習させることで、指定された回数だけ同じ処理を繰り返すといった複雑な制御構造を持つ情報処理過程を、入出力事例から近似的に学習できることを示している。

こうしたディープラーニングの研究の広がりは、データやタスクに適した特徴表現の学習の重要性を示している。その中で、大規模なネットワークの学習性能を向上させるための様々な工夫が生み出されている。また、Caffe、Chainer、TensorFlowなどの、層の数が多い複雑なニューラルネットワークモ

※6
二層からなる浅いニューラルネットワーク。

※7
入力を学習データとして、特徴量を抽出するニューラルネットワーク。

デルでも高速な学習と推論を簡単に実装するためのツールも、数多く作られて公開されている。

1.2.4 畳み込みニューラルネットワーク

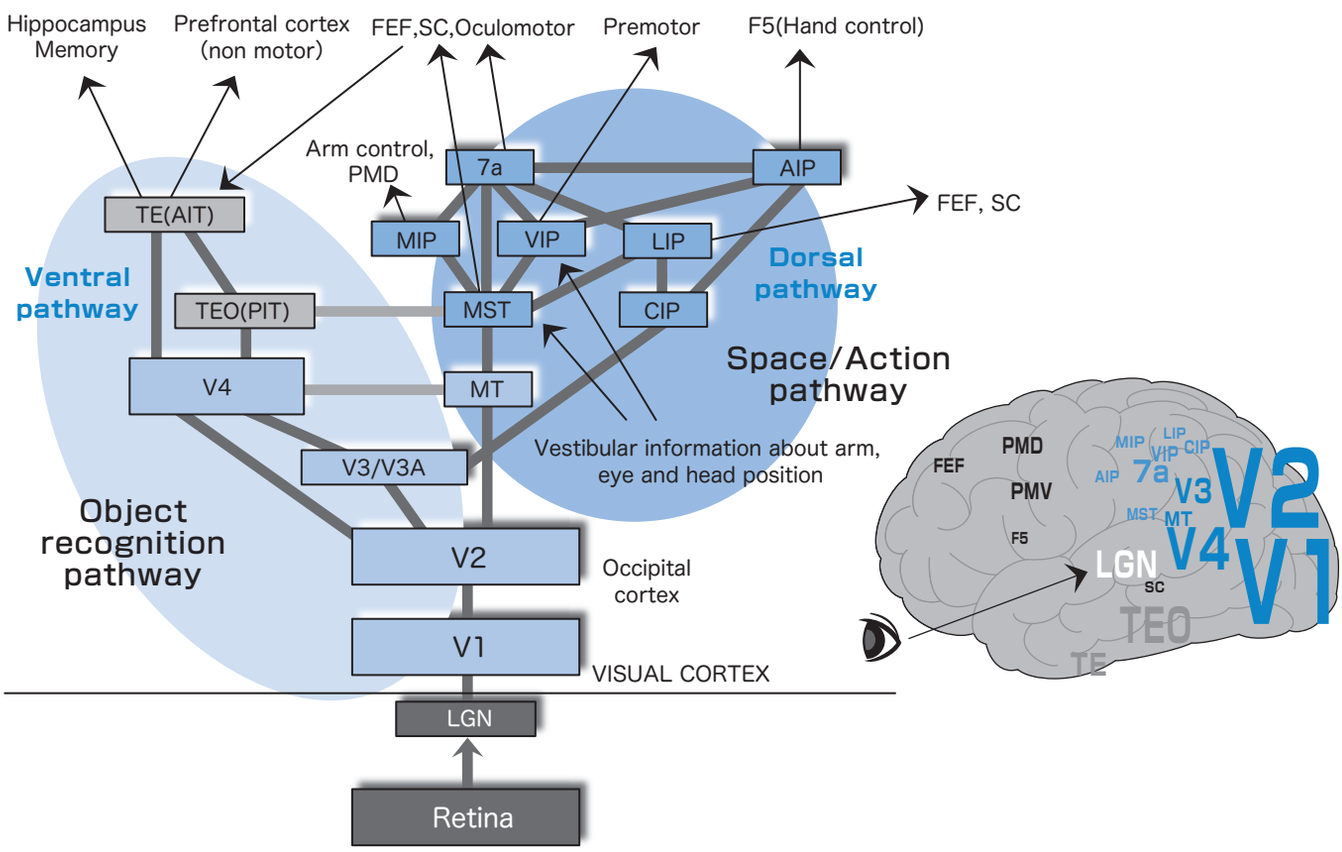
脳は、生理学的な特性によって“領野”（りょうや）と呼ばれる機能単位を有しており、視覚に関連する領野も様々な部位に分類できる。

図7は、脳の視覚野間の分類と関連性を模式的に表した図であり、視野中の形状に関する視覚信号は、腹側経路（V1野→V2野→V4野→IT（TE、TEO）野）と呼ばれる経路で処理される。

これらの領野のうち、比較的よく知られているのは初期視覚野と呼ばれるV1野の細胞群である。V1野の細胞は、視野中に受容野と呼ばれる情報処理範囲を持ち、一つの細胞の反応する範囲は非常に局所的である。これらの細胞は、受容野に示された入力刺激が、自身の“好み”に適合しないと発火状態にならない。この好みのことを選択特徴などと呼ぶ。

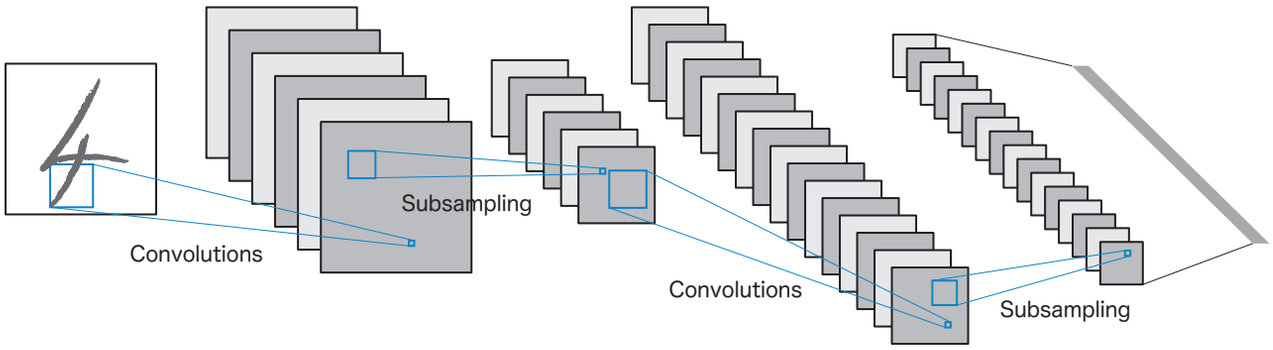
V1野細胞の受容野の大きさはおおよそ1度程度の大きさであり、抽出特徴としては、それほど複雑ではなく、画像中の「線分」や「エッジ」といった成分を抽出していると考えられている。更に、この選択特徴の性質の違いによって、初期視覚野の細胞は、単純型細胞と複雑型細胞に分類される。

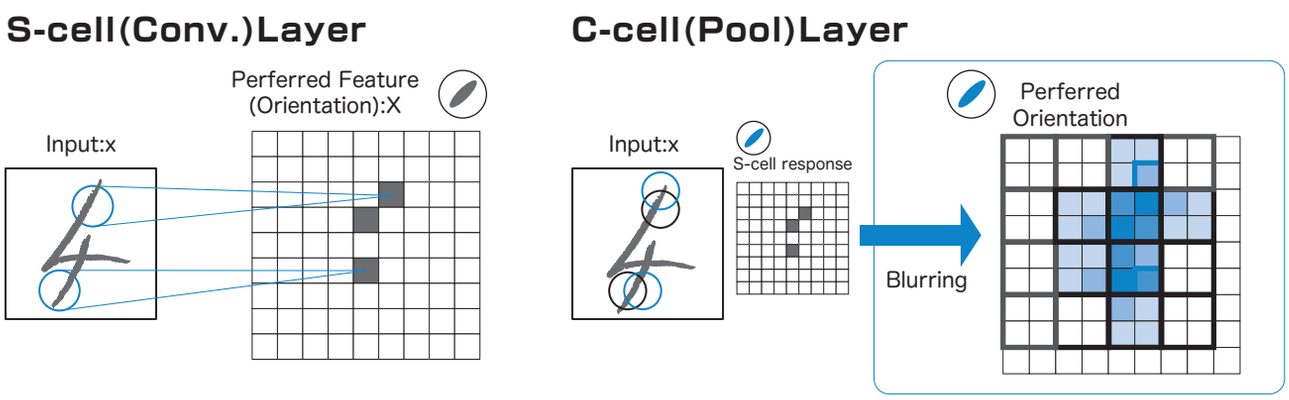
単純型細胞は、受容野内に提示される選択特徴の位置（位相と呼ぶ）に敏感で、受容野内に位置がずれて提示された選択特徴に対しては反応しない。一方、複雑型細胞も選択特徴を持つが、受容野内の提示位置の位置ずれに関しては寛容で、反応の低下はそれほど起こらない点が、単純型細胞との違いである。デイヴィッド・ヒューベル（David Hubel）氏とトルステン・ウィーセル（Torsten Wiesel）氏は、



■図7　視覚野のつながり[8][11]

※8
文献[11]より作成。





■図8　ネオコグニトロンアーキテクチャの概要

これらの単純型細胞と複雑型細胞とが、階層的に結合された関係であることを提案し（階層仮説）、現在のところ広く支持されている。

一方、高次視覚野であるV4野からIT野にかけては、耳や鼻といった顔の特徴的な部位に強く反応する細胞や、顔そのものに反応する細胞と抽象的な概念を符号化している細胞が観測される。更には、特定の人の画像のみならず個人名が書かれたテキスト画像にも同様に反応する、ある種の概念が符号化された細胞の存在も知られている。

これらの細胞では、受容野の大きさはかなり拡大され、IT野に至っては視野中のほぼ全域が受容野となる。すなわち、視野中のどこに物体や写真等を提示しても反応する細胞となっている。

福島邦彦氏は、これらの機能的な事実と生理学的な事実を基に、初期視覚野の細胞群のモデルとして以下のようなものを提案した（**図8**）。

まず、細胞が局所的な受容野を持つことを前提条件とし、このような細胞群を用いて画像全体を取り扱うために、同じ反応特性の受容野を持つ細胞をずらしながら2次元格子状に並べることとした。このように考えると、これらの細胞の計算は、工学分野で用いられる畳み込み（Convolution）演算として記述できる。この結果、畳み込み演算処理後の出力信号は、視野中のどこに選択特徴があったかを示す特徴マップとして表現される。

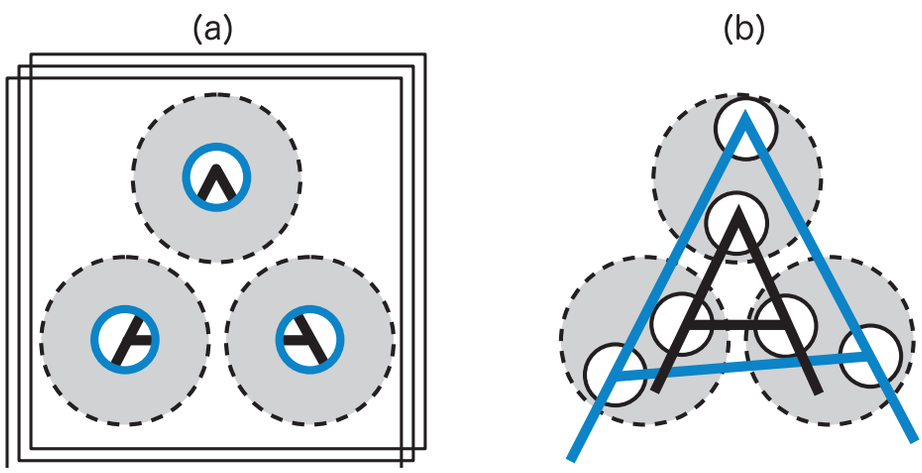
図8の左下の図では“斜め線”の特徴マップを表しており、特徴成分が入力のどこに出現しているのかを黒で示した。このように、単純型細胞は、入力層をいくつかの特徴成分に分解した特徴マップ群として表現する。この操作は視野中に含まれる信号から様々な特徴成分の抽出を行うことを意味する。一方、複雑型細胞は、特徴の位置ずれに寛容な細胞であり、選択特徴の位置（位相）がずれても反応する。位相がずれた位置に受容野を持つ細胞は、特徴マップとして考えた場合、位相ずれしていない位置に受容野を持つ細胞の周辺に存在するはずである。**図8**の右下の図では、斜め線の位相のずれた入力の受容野位置を薄青色の円で示している。すなわち、複雑型細胞の特徴マップは、単純型細胞の特徴マップを

空間的にぼかしたような特徴マップとして表現することができる。更に、複雑型細胞の特徴マップは、特徴の出現位置に対して冗長になるため、空間的なリサンプリング（プーリング）を行うことで空間的に情報圧縮を行うことが可能となる。

これらのネットワークアーキテクチャは「ネオコグニトロン」と呼ばれるニューラルネットワークモデルの基本構造となる。福島氏らのグループは、視覚野の高次機能を実現させるために、この基本アーキテクチャを外挿し、繰り返し表現の変換を行うことを提案している。このネオコグニトロンは、局所的な特徴を抽出し、空間プーリング位置ずれを許容しながら情報を圧縮することで、徐々に受容野を拡大し、情報を統合する脳の視覚野モデルとしてとらえることができる。福島氏らは、ネオコグニトロンを手書き文字認識などの問題に適用し、高度な識別器が実現できることを示してきた。また、ネオコグニトロンは、ニコス・ロゴテティス（Nikos Logothetis）氏らの示したIT野の細胞表現と類似した機能を実現し得ることが示唆されている[12]。

一方、画像処理分野におけるディープラーニングのモデルとして、既にデファクトスタンダードとしての地位を築きつつあるDeep CNN（DCNN）がある。福島のネオコグニトロンでは、クラスタリングに代表される教師なし学習を用いて学習することが提案されていたが、ネオコグニトロンの構造自身は、学習手法を限定するものではない。ニューヨーク大学のヤン・ルカン（Yann LeCun）氏らは、ネットワークアーキテクチャとして、このネオコグニトロンの基本構造を用い、学習手法として機械学習で用いられる誤差逆伝播法を用いたモデルを提案しており、これらの組合せが現在のCNNの基本となっている。2012年に、当時トロント大学（現在はGoogleに在籍）のアレックス・クリジェフスキー（Alex Krizhevsky）氏らが発表し、コンピュータビジョン分野に大きな衝撃をもたらした「AlexNet」や、2014年に発表された「VGG」（Visual Geometry Group）といったネットワークも、基本的にはネオコグニトロンのネットワークアーキテクチャを踏襲している。このように、現在のディープラーニングのオリジンの一部は、脳のモデルとしてとらえることができる。

図9にCNNの動作理解を行うための模式図を示す。図中の“A”という線画は、（b）の図に示した複数の“A”のように、異なる大きさと形状を持ち得るが、上端部の尖った形状とT字型の結合部の形状は、共通して持っている特徴ととらえられる。CNNの内部の畳み込み層（Convolution Layer）は、前述



■図9　CNNの動作原理の概要[9]

※9
「神経回路モデル「ネオコグニトロン」」『発明と発見のデジタル博物館』国立情報学研究所ウェブサイト <http://dbnst.nii.ac.jp/pro/detail/498> より改変（copyright 電子情報通信学会、許諾番号：17SB0039）

の単純型細胞群に対応し、これらの特徴を別々の特徴マップとして表現する。このため、**図9**の（a）のように、3種類の特徴マップで表現される。

一方、プーリング層（Pooling Layer）は複雑型細胞群に対応し、畳み込み層で得られた特徴の位置をぼかしたような特徴マップを出力する。このプーリング層の出力を受ける上位層では、**図9**の（a）の‘A’という文字に含まれる3種類の特徴のおおよその位置を組み合わせた特徴を、新たな高次特徴としてとらえる。このプーリング層のぼかし操作は、多少特徴の出現位置が変化しても許容する操作となる。

結果としてこれらの三つの特徴の組合せが揃っていれば、同じ特徴としてとらえる動作を行うため、変形などに強い特徴表現を構成できる（**図9**の（b））。このようにCNNでは、ネットワークが深い階層を取るほど、複雑な組合せ特徴を表現できるような構造になっている。

1.2.5 リカレントニューラルネットワーク

2015年の大規模画像認識チャレンジでは、「残渣ネット」（Residual Network; ResNet）（1.2.7項参照）の成績が人間の成績を上回った。2016年の結果も、前年の結果を踏襲しており、ResNetの変形によるショートカット付き多層化CNNモデルで、領域切り出しネットワーク（Region Proposal Network; RPN）、超高速領域CNNを複数合議制（アンサンブル）で精度向上を実現した。画像認識においてはCNNとそのアンサンブルで精度向上を目指す流れとなっている。このため多層CNNは画像認識、音声認識で一般的な手法となっていると考えられる。

そういったCNN（畳み込みニューラルネットワーク）は、ニューラルネットワークの層を一方通行で処理が行われる。

これに対して、再帰的な繋がりを持つネットワークとして「リカレントニューラルネットワーク」（RNN）に対する研究も進んでおり、自然言語や時系列データなどの連続性のあるデータに対して適用されている。「画素RNN」のように、静止画を右上から走査して系列情報処理モデルであるRNNモデルにより認識させるという手法も提案されている。この手法は時々刻々変化する中間層に何らかの表象が形成されることを仮定したモデルであり、一つの方向と言える。

2015年以降、「生成敵対ネットワーク」（Generative Adversarial Network; GAN）により画像を生成する手法が注目を集めた。この手法を用いれば、未知の画像を生成することができる。2014年以降にCNNの最終畳み込み層の結果をRNNへと接続したニューラルネットワークによる画像脚注生成技術と相まって、画像から言語、言語から画像、言語から音声や音楽など、異なる感覚様式を変換する手法への道が開かれた。このような流れはパターン認識技術の範疇を超え、AIと呼ぶにふさわしい領域へと一歩近づいたとみなして良いだろう。最終畳み込み層の上にソフトマックス層による認識層を使うのではなく、認識層の代わりに領域提案層、言語処理層、音韻処理層などを付け替えて、事前学習とファインチューニングを組み合わせることで学習時間の短縮を行うことが可能であるため、一般のユーザでも手軽に一般画像認識結果の恩恵を享受することができ、今日の熱狂につながったと言えよう。

RNNは系列予測、姿勢制御、自然言語処理への応用などが考えられてきた。近年、RNNの学習に関する勾配消失、勾配爆発問題を回避する手法が一般化したこと、大量のデータをGPUにより高速に処理できるようになったことを受け、性能を向上させてきた。2016年にはGoogleの自動翻訳サービスの精度が向上されたことが話題となったが、従来手法である統計的機械翻訳（Statistical Machine Translation; SMT）に対して、ニューラルネットワーク機械翻訳（Neural Machine Translation; NMT）が支配的になりつつある。この分野と画像処理において、任意の場所を選択的に処理する注意機構の導入は、画像と言語と領域は異なるものの数式は同一であり、脳内でも同じような機構が仮定で

きるのではないかと予想される。

アーケードゲームを解くために、認識技術にCNNを用い、高得点を得るために強化学習を用いる手法は、囲碁において世界チャンピオンレベルの強さを有するに至った。この枠組みは更に進歩しているが、かつてのチェスの木探索をブルートフォース[10]に行うのではなく、強化学習による一般的な解法を求めるアルゴリズムを用いている点が、従来からの特化型AIによる人間の専門家への挑戦という枠組みでは収まらない、一般的な解を模索する方法を提供しているように思われる。

以上のように、以前のブームをもたらした技術に比して、汎用性の高いアルゴリズムを用いていることがここでの特徴であり、今後更に発展するものと期待される。

1.2.6 表現学習

「表現学習」（representation learning）は、ディープラーニングを抽象化した概念である。機械学習の手法を構成する際に、有用な情報を抽出することができる、すなわちデータの特徴表現（あるいは、内部表現、素性（feature））を学ぶ学習の方法である。機械学習の性能は、データの表現に大きく依存しており、従来は人間の知識や職人技により、素性を構築することが広く行われてきた（素性工学（feature engineering）とも呼ばれる）。それを自動で学習するものが表現学習である。

例えば、曜日と天候から、店の売上を予測したいとする。それには、過去のデータを使って、曜日を表す変数x1と天候を表す変数x2から、お店の売上yを回帰する式を見つければよい。ところが、何を変数に置くかというのは、通常は問題の外にある。例えば、安売りをしているかどうかという変数x3を入れると、予測精度が大きく上がるかもしれない。

こうした特徴表現は、通常は人間の知識や職人技によって定義されるが、それによって機械学習の性能が大きく異なってしまう。あるいは、別の例では、画像に車が映っているかを判定したいとする。その際、ホイールが映っているか、ハンドルが映っているかなどを素性にすると良さそうであるが、画像から得られた画素の値という観測データ（あるいは生データ、RAWデータ）からそれらの素性をどう構成するのかというのは自明ではない。

特徴表現を学習する方法としては、従来から、様々なクラスタリング法、あるいは主成分分析（Principal Component Analysis; PCA）や独立成分分析（Independent Component Analysis; ICA）などの次元圧縮による手法がよく知られている。近年ではディープラーニングに注目が集まっているが、これも表現学習の一つである。ディープラーニングは、層の数が多いニューラルネットワークによって、観測データから本質的な情報を抽出した特徴表現を学習する[13]。

知的な情報処理とその学習における特徴表現の重要性は、AI、認知科学、機械学習、データ解析等の研究において古くから何度も指摘されてきた。デイヴィッド・マー（David Marr）氏は、いかなる計算機の計算処理も、計算理論、表現とアルゴリズム、ハードウェアという3階層から理解され得ると述べた。情報をどのように表現するかによって、アルゴリズムによる処理が大きく影響を受ける。

よい「表現」とは何かというのは、単純なようで難しい問いである。よい表現とは、何らかの意味で事象を抽象化したものであり、観測データの説明要因をとらえることで、一見自明ではない共通点をとらえることができるものである。ヨシュア・ベンジオ（Yoshua Bengio）氏は、よい表現に共通するものとして、世界に関する多くの一般的な事前知識（あるいは事前分布、prior）を挙げている。

代表的な事前知識に該当するものとして、

※10
すべての経路を総当たりで探索する。

（1）滑らかさ
（2）複数の説明要因
（3）説明要因の階層的構造
（4）半教師あり学習
（5）タスク間の共通要因
（6）多様体
（7）自然なクラスタ化
（8）時間的空間的一貫性
（9）スパース性（データの分布のまばらさ）
（10）要因の依存の単純性

などが挙げられる。例えば、よい表現とはスムーズな関数を用いるものであり、また時間的空間的一貫性を持っている。ディープラーニングのアプローチは、なかでも、（3）（5）（10）などに注目していることになる。逆に言えば、こうした事前知識を適切に活用することができるなら、表現学習は必ずしも層の数が多いニューラルネットワークの形をしていなくても良いということになる。

ディープラーニングで特徴的であるのは、簡単な関数の組み合わせで難しい関数を構成することである。通常、浅い構造よりも深い構造のほうがよりコンパクトに関数を表すことができる。そのことにより、素性の再利用と抽象的な概念（あるいは不変量）の獲得を可能にしている。ディープラーニングによる表現の獲得の例として、「深層信念ネットワーク」（Deep Belief Network）を用いて、インターネット上の動画から切り出した画像を入力し、猫の概念を生成したというGoogleの研究が有名である。また、抽象的な概念を頑健に獲得するには、設定の異なる複数のニューラルネットワークでの共通部分を見つけるとよいという最近の研究もある。

人間の知能においては、得られた抽象的な概念を、言語にマッピングし操作可能にしているところが特筆すべき点であろう。これに関連する有名な問題に、シンボルグラウンディング問題がある。AIにおける難問の一つであり[14]、スティーブン・ハーナッド（Steven Harnad）氏によって命名された。記号で指し示されるものを計算機がどのように認識するかという問題であり、概念に接続（グラウンド）されることなしには、記号処理が意味をなさないことを議論している。

更に遡れば、同様の議論は歴史的に古くからあり、スイスの言語哲学者であるフェルディナン・ド・ソシュール（Ferdinand de Saussure）氏は、記号内容と記号表記を、シニフィエ、シニフィアンと呼んだ[15]。また、イギリスの哲学者ジョン・ロック（John Locke）氏は、人間知性論のなかで、白紙の心の状態から概念が経験に由来して発生すると考えた[16]。

概念と言語のバインディングは、最近でも活発に研究が行われており[17][18]、またディープラーニングの文脈でも言語表現と画像特徴のアライメントを取るような研究も行われている。ディープラーニングを中心とする表現学習の方法の研究により、人間の持つ知性、特に言語を使った能力についての理解や、その工学的な応用も進んでいくことになるかもしれない。

表現学習においては、今のところディープラーニングのような深い構造を持ったニューラルネットワークを用いるアプローチが優勢であるが、必ずしもその構造が必然的なわけではない。（かなりいい線をいっているのは確かであるが）将来的には、今のニューラルネットワークと全く違う形での、より理論的な表現学習の手法が出現する可能性はあるだろう。

1.2.7 ディープラーニングの画像認識への応用

2012年に開催された一般物体認識のコンテスト「ILSVRC」（ImageNet Large Scale Visual Recognition Challenge）において、深い構造を持つCNN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。ここでは、ディープラーニングの画像認識への応用として、クラス分類、物体検出、物体セグメンテーション、画像キャプション生成、画像生成について述べる。

1.2.7.1 クラス分類

一般に、CNNは、層を沢山重ねて深い構造にすることで、より高い精度で物体を分類できるようになるが、その反面、パラメータ数が膨大となり学習が困難になる。そこで、深い層でも学習がうまくいく枠組みとして、「ResNet」が提案されている。ResNetは、出力を入力と入力からの差分の和でモデル化したネットワークである。この構造によって、上層からの誤差が下層まで伝播するようになり、1000層といったかなり深い構造でも適切に学習が可能となった。

ResNetは、ILSVRC2015の様々な部門において、トップの成績を獲得した。このときの物体クラス分類課題における、上位5位までに正解が含まれないエラー率は、3.57%であった。一方、ILSVRC2016における同部門の1位のエラー率は2.99%、2位が3.03%、3位が3.04%であり、2015年と比較して分類性能はほとんど伸びていない。また、ILSVRC2016のトップの手法は既存技術の組合せであり、物体クラス分類において、この1年間はインパクトのあるトピックが出ていない状況にある。

1.2.7.2 物体検出

物体検出とは、画像内の物体を取り囲むボックスを推定するタスクである。物体検出においても、ディープラーニングを利用して検出精度の向上が実現されている。ディープラーニングを利用した物体検出の例として「R-CNN」（Regions with CNN）がある。

R-CNNでは、選択的検索法から得られる物体領域候補内の画像を、事前に学習しておいたCNNに入力し、この領域の画像特徴を抽出する。次に、線形サポートベクトルマシンに抽出された画像特徴を入力し、領域の物体クラスを予測する。R-CNNによって高い検出性能は得られるが、R-CNNは物体領域候補の数だけCNNの順向き伝播の計算が必要である。また、R-CNNは物体検出ネットワークとは別のモジュールで物体領域候補群を計算する必要があった。そこで、物体検出ネットワークと共通の特徴マップから領域候補群を提案するネットワークを作り、この物体検出ネットワークと領域提案ネットワークを統合した「Faster R-CNN」が提案されている。

上記手法では、推定された物体領域候補に物体クラス分類手法を適用することで、画像内の物体の検出を行っていた。物体検出を回帰問題ととらえてモデル化することにより、物体領域を提案するネットワークを不要とし、一つのネットワークで実現できるアルゴリズムも提案されている。また、検出性能を直接的かつEnd to Endで最適化可能であり、検出自体も高速に実行できる。

1.2.7.3 物体セグメンテーション

物体セグメンテーションは、物体を取り囲むボックスではなく、対象物体と背景を境界まで詳細に切り分けるタスクである。ディープラーニングを利用した代表的な手法として「FCN」（Fully Convolutional Network）がある。クラス分類のネットワークは前段が畳み込み層、後段が全結合層となっている。カーネルサイズを入力特徴マップのサイズと同じにすれば、全結合層を畳み込み層とみなすことができる。そこで、クラス分類のネットワークの全結合層を畳み込み層に置き換えることで、どの領域に

何がありそうかを表現した分類マップを得ることができる。しかしながら、このままではプーリングの影響で分類マップの解像度が低いため、分類マップを入力画像サイズにアップサンプリングすることで最終的な物体セグメンテーション結果を得る。

上記手法は、ピクセルレベルでセグメンテーションを行うため、意味レベルでの物体セグメンテーションには適切ではない。例えば、複数のリンゴが隣接して置かれた場合、リンゴ同士を切り分けることは困難である。そこで、各物体を分けつつ物体を背景から切り出す手法として「MNC」（Multi-task Network Cascade）などが提案されている。

1.2.7.4 画像キャプション生成

現在の潮流として、画像と自然言語処理の融合分野がある。この融合分野のタスクとして、画像から「赤い服を着た女性が街中で電話をしている」のような、自然言語で記述された画像キャプションを生成することが挙げられる。基本的なキャプション生成の流れは、画像をCNNに入力し、CNNから得られた画像特徴を、時系列を扱えるネットワークであるLSTMに入力する。LSTMは内部記憶を持っており、事前に生成した単語を考慮しながら、単語を次々と生成していき、最終的な文章を作り出す。

1.2.7.5 画像生成

画像生成も注目を浴びている技術である。2015年にGoogleが「Deep Dream」と呼ばれるシステムを開発し、大きな話題となった。Deep Dreamは、通常の画像を夢に出てくるような神秘的な画像に変換するシステムである。

また、Googleは「Deep Style」と呼ばれる、入力画像の画風（例えば、ゴッホなど）を変換するシステムを開発し、AIが製作する芸術作品としてメディア等で取り上げられている。

いま最も利用されている画像生成手法は、生成敵対ネットワークを利用している。このネットワークは、画像生成器と画像識別器から構成されており、画像生成器は分類器を騙すような画像を生成し、識別器は生成器から生成された画像と本物の画像とを分類するようにそれぞれ学習する。このように競合して学習することで、生成器は適切な画像を生成することが可能となる。

1.2.8 ディープラーニングの音声認識への応用

音声認識においてニューラルネットワークを用いる研究は、1990年代初頭に活発に行われたが、その後は混合正規分布（Gaussian Mixture Model; GMM）に基づく隠れマルコフモデル（Hidden Markov Model; HMM）が一般的となった。大規模なデータを収集して、多数のテンプレート（sum of experts）を用意すれば性能がいくらでもよくなると考えられていた。これに対して2010年頃にヒントン氏らが、多段のネットワーク（product of experts）を学習するディープラーニングにより、一般的な音素認識タスク（Texas Instruments Massachusetts and Institute of Technology; TIMIT）で驚くべき性能を挙げた。その後MicrosoftやIBM、Googleなどの研究者らにより、種々の大語彙連続音声認識でも大きな改善が得られることが示された。音声認識は、ディープラーニングが最初に成功を収めたタスクの一つである。

現在、世の中で一般的な音声認識システムの構成を**図10**に示す。この音素状態の認識において、GMMの代わりに深層ニューラルネットワーク（Deep Neural Network; DNN）を用いているのが眼目である。その後の展開を含めて、**図10**の各要素におけるディープラーニングの導入について以下に述べる。

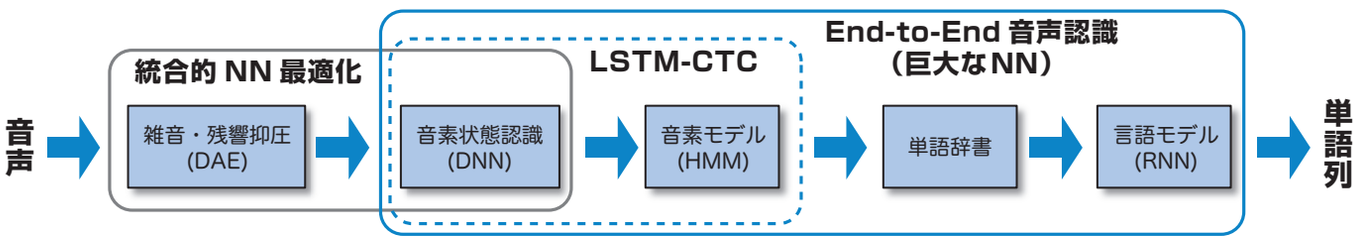
1.2.8.1 音素モデルにおけるディープラーニング

前記のとおり、DNNは音響特徴量（音声の周波数スペクトル）を入力として、音素状態（音素を細分化した数千クラス）のカテゴリに分類する。これは基本的には教師付き学習であるが、そのためには事前に音声データを音素状態に区分化する必要がある。これはGMM-HMMなどを用いて自動的に行われる。音素や単語の時系列のモデル化には、HMMが依然として用いられている。

これに対して近年、HMMを一切用いずに、RNNの一種であるLSTMを用いて、音響特徴量の系列から音素や文字などの系列に直接変換するCTC（Connectionist Temporal Classification）という方式が注目を集め、従来のDNN-HMMの性能に近づきつつある。

1.2.8.2 言語モデルにおけるディープラーニング

言語モデルは単語系列仮説の尤度（ゆうど）[11]を評価するもので、従来はN-gramモデルが一般的に用いられてきた。これに対して近年、RNNを用いたモデルも導入・併用され、認識率の改善を実現している。



■図10　典型的な音声認識システムの構成とディープラーニングの導入

1.2.8.3 雑音・残響抑圧におけるディープラーニング

自動車内や家庭内の機器やロボットの音声認識においては、雑音や残響の影響が大きな問題となり、音響特徴量の計算においてこれらの影響を抑圧する処理が必要となる。従来は、線形フィルタや統計的なモデルが用いられてきたが、近年はデノイジングオートエンコーダ（Denoising Auto Encoder; DAE）などのディープラーニングが用いられている。また、複数のマイクロフォンを用いて音声を強調する場合でも、雑音の分離などにDNNが用いられる。

更に、この雑音・残響抑圧のDNNと音素状態認識のDNNを連結して巨大なニューラルネットワークを構築し、統合的に最適化することも検討されている。

1.2.8.4 End to Endのモデル化

前記のCTCを含めて、音響特徴量から音素、文字列、更には単語列に直接変換するアプローチをEnd to Endモデル[12]と呼ぶ。単語辞書や言語モデルも含めて、統合的にディープラーニングの枠組みで構築することも検討されている。ただし、日本語や英語などの主要言語では、テキストのみのデータが圧倒的に多いので、言語モデルを単独に学習したほうがよいのは明らかである。

※11
統計学において、ある条件から結果を得た場合に、結果から逆に条件を推測する際の尤もらしさを表す数値。

※12
End to End Learningは深層学習の重要な方法論の一つ。従来は、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が必要であったが、深層学習により、入力から出力までを一つのネットワークとして表現することで全体のネットワークを一気に学習できるようになった。

1.2.8.5 音声合成におけるディープラーニング

音声認識の逆過程である音声合成においても、ディープラーニングの導入が活発に行われている。2016年に発表された「WaveNet」では、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している。

ニューラルネットワークはブラックボックスであるが、可塑性が高いので、複数言語のモデルを統合的に学習したり、画像情報と組み合わせたり、様々な展開が行われている。また、音声認識の典型的な後処理である対話や翻訳においてもディープラーニングが導入されているので、それらとの密な結合も今後考えられる。

1.2.9 ディープラーニングの芸術への応用

一般的には極めて人間的な行為と考えられている創造性が必要な絵画や音楽等の芸術分野においてもAIの応用が始まっている。

芸術への応用を目指す研究の端緒の一つは、画像を生成することに関する研究がいくつか発表されたことである。2015年、GoogleがAIを用いた画像処理アルゴリズムとして「Deep Dream」を公開した。Deep Dreamは、学習済みのCNNの内部がどのようになっているかを知るために開発された手法である。学習済みのCNNに対し、指定した画像を入力し、写っている物体を例えば「犬」であると認識したとする。その場合、「犬」という判定結果を強調するように元の画像を少し変化させる。これを繰り返すことにより、画像全体を変化させていくものである。このように、本来は学習済みのCNNの中を知るために開発されたアルゴリズムであるが、生成されるパターンがサイケデリックな画像となるため、一般にも注目を集めた。

また、同年、Googleはディープラーニングを用いた画像生成手法として「DRAW」（Deep Recurrent Attentive Writer）を発表している。DRAWアルゴリズムは、生成モデルとして「VAE」（Variational Auto Encoder）を用い、さらに画像中の各部分に注意を向けながらRNNで反復的に画像を生

■図11　BEGANにより生成された顔画像の例[13]

※13
David Berthelot et al., “BEGAN: Boundary Equilibrium Generative Adversarial Networks.” Cornell University Library Website <https://arxiv.org/pdf/1703.10717.pdf>

成するモデルであり、数字の画像を生成することに成功した。更に、2016年には、自然言語で書かれた文に対応した画像を生成する「AlignDRAW」がトロント大学により発表されている。ぼやけた画像ではあるものの、「A stop sign is flying in blue skies」（停止標識が青空を飛んでいる）、「A toilet seat sits open in the grass field」（芝生のなかに便座が開いている）など、通常存在しない状況に対応した画像も生成可能であることを示した。また、Googleは「BEGAN」（Boundary Equilibrium Generative Adversarial Network）という生成モデルを用いて、自然な人物の顔画像の生成に成功している（**図11**）。

Deep Dreamと同時期に、画像や絵をディープラーニングで生成する技術を芸術の分野で応用しようという研究が出始めた。2015年、ドイツのチュービンゲン大学の研究者らによって、「A Neural Algorithm of Artistic Style」と呼ばれるアルゴリズムが開発された。このアルゴリズムでは、葛飾北斎やゴッホらの画風を特徴量（「スタイル」と呼ばれる）として学習し、任意の画像にスタイルを重ねて出力することで、北斎風やゴッホ風のように変換できるようにしたものである。

2016年には生成モデルの「DCGAN」（Deep Convolutional Generative Adversarial Network）を用いて同様のことが可能であることが示されている[14]。

また、デルフト工科大学（オランダ）とMicrosoftは、17世紀の画家であるレンブラントの絵画の題材や筆づかい、色合いといった作品が持つ特徴をディープラーニングにより分析して、3Dプリンタによって再現するプロジェクト「The Next Rembrandt」を推進しており、2016年にその成果を公開した[15]。レンブラントの全作品を3Dスキャンでデジタル化し、オランダのマウリッツハイス美術館とレンブラントハイス美術館の専門家の協力を得て、ディープラーニングによって作品の特徴が分析された。

音楽分野への応用も始まっている。ソニーコンピュータサイエンス研究所（Sony CSL）は2016年、AIを使ってリードシート[16]を登録したデータベースから音楽スタイルを学習し、学習したスタイルを用いて自動的に作曲することを目指すプロジェクト「Flow Machines」を発表した[17]。このアルゴリズムを用いて、例えばビートルズ風のスタイルを指定すると、実際に作曲が可能であることを示した[18]。また、プリンストン大学のキム・チソン（Ji-Sung Kim）氏が2016年に開発した「deepjazz」[19]は、2層のLSTMを用いてジャズの楽曲を学習し、ジャズを生成できるようにした。

そのほか、カリフォルニア大学サンディエゴ校のクリス・ドナヒュー（Chris Donahue）氏らが、2017年に任意の音楽を入力するとゲームセンターの音楽ダンスゲームである「ダンスダンスレボリューション」のステップを生成するアルゴリズムである「Dance Dance Convolution」を発表した。Dance Dance Convolutionでは、入力された音楽の特徴的なタイミング（ステップを踏むべき点）を音楽のスペクトルから学習し、難易度を指定するとLSTMを用いてステップを生成する。

また、AIで小説を生成しようという試みとして、公立はこだて未来大学の松原仁氏らによる、ショー

※14
「ディープラーニングによる傑作：人工知能の画期的なスタイルを紹介するキャンバスがGTCに登場」NVIDIAウェブサイト <https://blogs.nvidia.co.jp/2016/04/11/artificial-intelligence/>
※15
「人工知能が描いた「レンブラントの新作」」WIREDウェブサイト <http://wired.jp/2016/04/14/new-rembrandt-painting/>
※16
ポピュラー音楽の歌や、ジャズの曲を楽譜にするときによく使われる、基本的な部分のみを取り出して紙などに書きあらわす記譜法。
※17
Flow Machinesは、欧州研究評議会（European Research Council; ERC）からの資金提供を受けて開発されている。Flow Machines Website <http://www.flow-machines.com/>
※18
「世界初の人工知能が作ったポップソング「Daddy's Car」と「Mr Shadow」がYouTubeで公開中」Gigazineウェブサイト <http://gigazine.net/news/20160924-daddys-car-ai-song/>
※19
deepjazz Website <https://deepjazz.io/>
※20
「きまぐれ人工知能プロジェクト作家ですのよ」公立はこだて未来大学ウェブサイト
<https://www.fun.ac.jp/~kimagure_ai/index.html>

トショートを創作させることを目指すプロジェクトがある[20]。2016年に行われたショートショート分野の新人賞である第3回星新一賞[21]では、作成時にAIを利用した2作品が一次審査を突破した。現状では、作品作成時におけるAIの利用割合は部分的であり、近い将来に完全にAIでの執筆が可能になることは困難と考えられる。ただし、その前段階として、人間が小説を執筆する際の支援に利用できるようになる可能性がある[22]。

1.2.10 ディープラーニングの実現技術

1.2.10.1 ディープラーニングのソフトウェア

ディープラーニングはソフトウェアフレームワークを利用して実装するのが一般的である。多層のニューラルネットワークモデルを定義し、メモリ上に対応する計算グラフを構築し、データを用いて学習・予測を実行するのがフレームワークの役割だが、重要なのはネットワーク記述方法とその柔軟性である。

ネットワークの記述方法の一つ目は、設定ファイルによる記述であり、「Caffe」や「CNTK」（Microsoft Cognitive Toolkit）などがこの方式を採用している。ユーザはデータに適用する関数の種類とその設定を順に並べたテキスト形式の設定ファイルを用意し、これをフレームワークが読み込んでネットワークを構築する。ネットワークの定義自体がテキストデータとなるために可搬性が高く、実システムへの組込みが容易になる一方、ループ構造を持つRNNなどのように、複雑になると人手でネットワーク構造を記述することは難しい。

二つ目はプログラムによる記述であり、「TensorFlow」や「Chainer」など、多くのフレームワークが採用している。ループ構造などもプログラムであれば簡単に記述できる。一方でネットワークの定義がソースコードの形で与えられるため、可搬性はそのプログラミング言語に依存する。

学習を行う際は、ネットワークの定義からモデルをメモリ内に構築したあと、訓練データを入力して順方向計算と誤差逆伝播による勾配計算を行って、パラメータ更新を行うのが一般のフレームワークのアプローチである。この方法では、学習を行う前にモデルが固定化されるため、その際にネットワークの性能面の最適化を行える。ただし、柔軟さに欠けるため、学習時に確率的に構造が変化する動的なニューラルネットワークを扱うのが得意ではなく、特別な計算コンポーネントを導入する必要がある。

一方、プログラム中に直接順伝播の処理を書き、毎回その処理を呼び出すことでネットワークの構築と学習を同時に行う柔軟なアプローチも存在する。これはChainerや「PyTorch」が採用しており、Define by Runや動的計算グラフとも呼ばれる。プログラムのデバッグや性能のプロファイリングは容易であるが、計算上のオーバーヘッドは増える傾向がある。だが、近年アルゴリズム提案が増えている動的なニューラルネットワークを、プログラムとして直感的に記述できるため、その重要性は増している。

1.2.10.2 ディープラーニングの実装

（1）行列演算による実装

神経回路のアナロジーでは、多数のユニットを結ぶ重み付き有向辺（矢印の付いた辺）で表現されるニューラルネットワークだが、計算を効率的に行うために、行列演算と非線形関数の組合せで書かれることが多い。行列演算は科学計算一般で用いられるため、高度に最適化された既存の実装が利用可能で

※21
「第3回　日経「星新一賞」」日経星新一賞ウェブサイト
<http://hoshiaward.nikkei.co.jp/no3_result/index.html>

※22
「AI、小説の大海原に乗り出す 作家誕生の日はいつ？」NIKKEI STYLE ウェブサイト
<http://style.nikkei.com/article/DGXKZO03732040X10C16A6BC8000>

ある。画像認識などが応用先にあることを考えると、ベクトルと行列だけでなく三つ以上の軸を持つテンソルも扱えることが望ましい。

(2) 勾配の計算手法

行列演算と非線形関数で記述されたニューラルネットワークは、それらを合成した巨大な関数と見なせる。これを勾配法によって最適化する場合、各辺の重みやバイアス項に関する勾配を求める必要がある。たとえ巨大なネットワークであっても、合成関数として与えられているので、その微分は連鎖律を用いて各関数のヤコビ行列の積として展開できる。

ディープラーニングで最も一般的な勾配計算法は誤差逆伝播法で、連鎖律によって展開された勾配を、出力に近い関数のヤコビアンから逆順に求めるために、こう呼ばれる。合成関数の勾配計算には自動微分（automatic differentiation）と呼ばれる数値計算アルゴリズムが用いられる。

CNNなどの単純なフィードフォワードネットワークだけでなく、RNNでも計算手順を時間方向に展開することで、誤差逆伝播法を用いることができる。これを通時的逆伝播（backpropagation through time）と呼ぶ。途中状態を全て保存しなければならないため、実際には長い系列では逆伝播を打ち切るヒューリスティクスが用いられる。

(3) 最適化ルーチンと効率化

ニューラルネットワークの勾配を計算したあとは、確率的勾配法に基づく最適化ルーチンを実行することで、パラメータを更新する。最適化アルゴリズムとしては単純なSGD（Stochastic Gradient Descent）に加え、RMSProp、Adamなどがよく用いられる。一般には全てのパラメータについて同一のものを選択する。

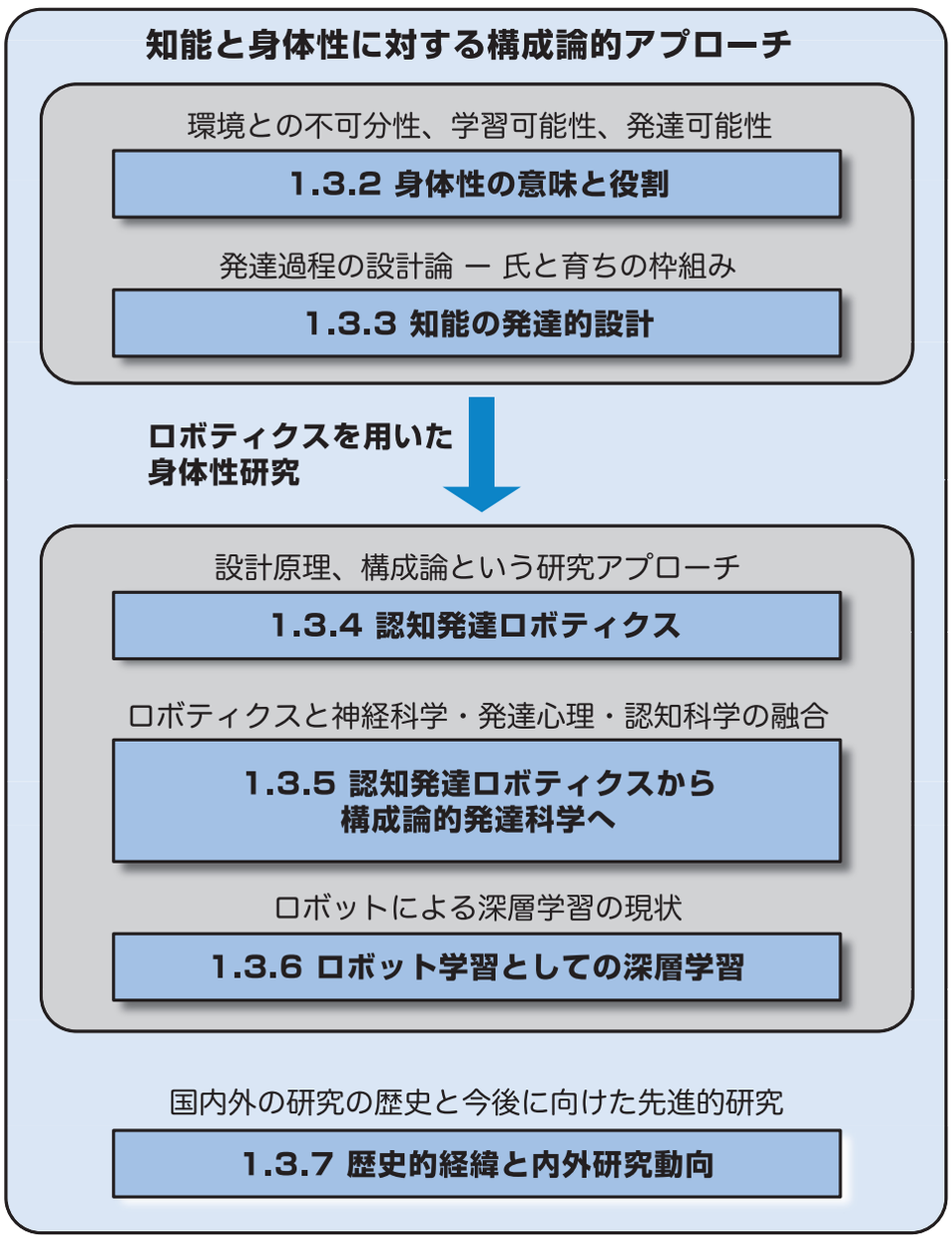
また大量の訓練データを高速に処理するため、オンライン学習で複数の学習サンプルからまとめて勾配を計算するミニバッチと、その並列計算が併用されることが多い。特にNVIDIAの汎用GPUを用いた大幅な高速化は、大規模データセットを用いて複雑なニューラルネットワークを現実的な時間で学習するために必須である。

参考文献

[1] 銅谷賢治ほか「小脳、大脳基底核、大脳皮質の機能分化と統合」『科学』vol.70 No.9, pp.740-749.
[2] 浅田稔・國吉康夫『ロボットインテリジェンス』岩波書店.
[3] 高橋泰岳・浅田稔「実ロボットによる行動学習のための状態空間の漸次的構成」『日本ロボット学会誌』vol.17 No.1, pp.118-124.
[4] 保田俊行・大倉和博「強化学習の最近の発展《第8回》連続空間における強化学習によるマルチロボットシステムの協調行動獲得」『計測と制御』vol.52 No.7, pp.648-655.
[5] J. H. Connel and S. Mahadevan, “Introduction to robot learning,” Robot Learning, Kluwer Academic Publishers, pp1-17.
[6] 川人光男ほか「多重順逆対モデル（モザイク）その情報処理と可能性」『科学』vol.70 No.11, pp.1009-1018.
[7] 高橋泰岳・浅田稔「複数の学習器の階層的構築による行動獲得」『日本ロボット学会誌』vol.18 No.7, pp.1040-1046.
[8] 牧野貴樹ほか『これからの強化学習』森北出版.
[9] 浅田稔ほか「内発的動機付けによるエージェントの学習と発達」『これからの強化学習』森北出版.
[10] 麻生英樹ほか『深層学習－Deep Learning－』近代科学社.
[11] N. Kruger et al., “Deep Hierarchies in the Primate Visual Cortex: What Can We Learn for Computer Vision?” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.35 No.8, pp.1847-1841.
[12] 吉塚武治ほか「ネオコグニトロンによる視覚腹側経路のモデル化」『日本神経回路学会誌』vol.14 No.4, pp.266-272.
[13] 麻生英樹「多層ニューラルネットワークによる深層表現の学習」『人工知能学会誌』vol.28 No.4, pp.649-659.
[14] 松原仁「一般化フレーム問題の提唱」『人工知能になぜ哲学が必要か』哲学書房.
[15] フェルディナン・ド・ソシュール（影浦峡・田中久美子訳）『ソシュール一般言語学講義：コンスタンタンのノート』東京大学出版会.
[16] ジョン・ロック（大槻春彦訳）『人間知性論１』岩波書店.
[17] 谷口忠大『記号創発ロボティクス 知能のメカニズム入門』講談社.
[18] 谷淳「ロボットで「科学」する記号の問題」『日本ロボット学会誌』vol.28 No.4, pp.522-531.

1.3 身体性と知能の発達

1.3.1 総論



■図12　本節の構成

人の知能に関する研究は、従来、知能がいかに動作して、その機能を実現しているかについて、分析的かつ明示的な設計に基づいて構成しようとする方法論で行われてきた。このような方法論は、説明原理に基づく研究と呼ばれる。それに対して、構成論的アプローチでは、比較的単純な身体を構成し、環境と身体の相互作用によって知的な振る舞いがボトムアップに生成されるかどうかを調べる。いわば、知能を創ってみることで研究するという方法論である。知能の設計原理に着目した研究ということもできる。

したがって、構成論的アプローチでは、比較的単純な構成を初期条件として用意し、環境と相互作用しながら、自ら学習して成長していく知能の主体を創ることを目指す。このような知能の見方においては、環境との相互作用の媒体となる身体がクローズアップされ、「身体性」と呼ばれる重要な研究テーマとなる（1.3.2項参照）。

主体から見れば、身体を通してしか環境への働きかけができず、環境からの情報のフィードバックも身体からしか受け取れない。知能の主体は、このような認知的な枠組みの中で多様な行動を行い、学習していくのである。具体的な研究手法としては、身体性について、人間を始めとする生物システムと人工システムに共有可能な性質としての定義を与え、その役割や期待される機能を明示するとともに、脳神経系、筋骨格系、体表面の各部がミクロレベルからマクロに渡って実現可能かについて考察するという手法が採られる。

また、構成論的アプローチは、知能の時間的な成長を実現しようとするものであるため、知能がどのように発達し得るかという観点が重要となる。赤ちゃんの発達過程の大きな特徴の一つは、創発的であることである。創発とは、形成される秩序の形態を明示的に指定するのではなく、比較的単純な下位の法則から自発的に秩序が形成されることを指すが、赤ちゃんは、環境の中で能動的に探索を行う中で、自律的に概念形成を実現する。

その過程は、外部から教師データを与えられずに、自律的に概念が獲得されるという意味で創発的である。更に、比較的初期の内的世界の発達でキーとなる機構に、「ミラーニューロンシステム」がある。ミラーニューロンシステムは、運動の生成と観測を結ぶことがポイントであり、そのことが、自己と他者の運動の同一性の理解から、他者意図理解などに発展する可能性を持つ（1.3.3項参照）。

以上のような、知能の自律的な発達を可能とする身体を創ってみることによって研究するという構成論的アプローチを、具体的にロボットを用いて研究する手法として、認知発達ロボティクスを導入する（1.3.4項参照）。その基本的な考え方としては、従来の説明原理に基づくロボティクスとは異なり、ロボット自らが、学習過程を通じて知能を発達させる点が挙げられる。この設計過程の中で、脳・神経科学、発達心理、認知科学などの既存分野と連携し、人間の認知発達の未解明の仕組みにも迫る意味や価値を示す。

また、認知発達ロボティクスの枠組みの中で、特に言語を始めとする記号自体がどのように創発したかを問う記号創発ロボティクスを紹介する。ここでの基本的な考え方は、人工知能の設計を記号システムありきで始めずに、記号システムが環境適応を通して形成されるプロセスを機械学習により表現し、言語獲得や社会における言語の形成を、ロボットを用いて構成論的に理解しようとすることである。

記号創発ロボティクスやそれを含む認知発達ロボティクスの考え方を、より発展させたものとして構成的発達科学を紹介する（1.3.5項参照）。知能設計の研究には、機械学習やロボティクスだけでなく、神経科学、発達心理学、認知科学など周辺の幅広い諸科学が深く関係する。これらの諸科学を発達の研究のために総合する学問が構成的発達科学である。例としては、自閉症などの非定型発達の当事者による研究や、共感発達を例に自他認知の課題を扱う研究が、構成論的アプローチをベースとして行われている。

ディープラーニングは、ほとんど同様の枠組みで多様なモダリティ[1]を扱えることから、ロボットの行動学習に対しても非常に有効であると考えられる。また、ディープラーニングのロボットの行動学習への応用例を紹介し、優れた学習結果、並びに今後の課題を示す（1.3.6項参照）。

最後に、歴史的経緯と内外の研究動向を述べ、身体性と知能発達の今後の課題を示す（1.3.7項参照）。

※1
視覚、聴覚、触覚等の感覚の種類のこと。

1.3.2 身体性の意味と役割

1.3.2.1 知能にとっての身体性とは？

前述のように、知能が「経験、学習、発達」するためには、身体を持っているということ、すなわち「身体性」が重要である。身体性は、「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」と規定される[1]。そのような身体性は、以下のような性質をもたらす。

1. 不可分性：
   様々な環境やその変動及び、知能の主体自身の内部状態を感知できる感覚能力、環境に働きかける多様な運動能力、それらを結ぶ情報処理能力は不可分であり、密に結合していること。
2. 学習可能性：
   限られた資源（感覚の種類や能力、運動能力）や処理能力の範囲で目的を達成するために、知覚・運動空間の関係の経験（環境との相互作用）を通じて学習できること。
3. 発達可能性：
   達成すべき目標や環境の複雑さの増大に対して、適応的に対処できるように、学習結果の経時的発展（発達）を可能にすること。

これらのことを示す興味ある実験として、生後2週間の2匹の仔猫の生理実験がある。1匹は自ら歩行し、もう1匹は自ら歩行できない状態で、同じ視覚入力を与えた場合、前者が正常な奥行き知覚を構成できるのに対し、後者は正しく視覚入力を解釈できず、正常な奥行き知覚が構成されないと言われている。このことは、感覚入力を正しく理解するには、自らの能動的な運動入力が不可欠であること（不可分性）、それらの関係を学習できること（学習可能性）が大事であることを示している。更に、自ら歩行していない仔猫でも、歩けるような環境を与えれば、正しい知覚が構成される（発達可能性）。

身体を、感覚・運動・認知を支える物理的基盤と考えると、身体の物理的構造による拘束（形態）だけでなく、感覚器、運動器、内臓など、どのレベルまで生物学的な意味合いで、その内部構造を模擬するかが、議論されている。

以下では、この身体性を構成する各部についての現状と課題について探る。構成的手法に基づく研究でこれまで扱ってきた身体には、脳神経－感覚器－筋骨格－体表面系の一部しか含まれておらず、現状では、消化器系や循環器系、呼吸器系などは、明示的には含まれていない。下記では、脳神経系、筋骨格系、体表面に関する研究の現状を述べる。

1.3.2.2 脳神経系

現在までの脳神経系に関する構成的研究では、運動野と感覚野のみのごく一部を扱っているか、他の脳部位を想定していても、明確な対応付けが困難な場合が多い。更に、発達の視点も考慮に入れると、対応問題が難しいというよりも、発達の過程で対応が変化するため、明示的な対応は不適切となる可能性も高い。

例えば、言語発達の初期においては、左脳より右脳の障害のダメージが大きいことが知られている。このことは、成人において最終的な言語野と呼ばれる部分が、最初から中心的な役割を果たしているのではなく、発達の初期段階では、異なる部位が関わっているためと考えられる。また、意識的な注意の機構は、視覚情報の顕著性などによるボトムアップ的なものから、様々なタスクを遂行する上で必要なトップダウンの注意に発達し、関連する脳部位は、後方から前方へ移動することが知られている。

個体発達をメインにしたボトムアップ的なアプローチでは、最小実装から始めて、徐々に機能と構造を複雑化するアプローチに加え、他者を含めた環境、特に養育者との相互作用を主体としたモデル化に準じた脳神経系の設計原理が必要である。さらに、ミラーニューロンシステム（後述）を始めとするメカニズムの獲得も重要な課題となる。

1.3.2.3 筋骨格系

筋骨格系は、人間を始めとする動物の運動を生成する身体の基本構造である。これは、従来のロボットではジョイントリンク構造に相当するが、大きな違いは、アクチュエータ[2]として動物では筋肉が、ロボットでは主に電動モータが利用されている点である。電動モータは、制御が容易であるなどの観点から、アクチュエータの代表であり、様々に利用されている。ただし、制御対象と制御手法を区別し、制御手法を駆使することで、様々な動きを実現可能であるが、撃力を伴う衝突やトルクや速度の極めて大きな変化を含む激しい運動は非常に困難である。

これに対して、動物では、筋骨格系身体を効率的に利用して、跳躍・着地、打撃（パンチ、キック）、投擲（ピッチング、砲丸投げ）などの瞬発的な動作を実現可能である。また、動物の筋骨格の構造としては、一つの関節に対し複数の筋肉が、また一つの筋肉が複数の関節にまたがって張りめぐらされ、複雑な構造となっている[2]。そのため個々の関節の個別制御により動きを実現するのではなく、身体全体として、環境と相互作用するなかで動きが自発的に生成され、運動を獲得する。一見、不都合に見えるが、逆に超多自由度ロボットにおける自由度拘束問題[3]の解決策とも言える。

このような生物にならう筋骨格系の人工筋として、「McKibben型空気圧アクチュエータ」[4]が注目されている。これを用いた跳躍ロボットが開発され、すでに動的な運動を実現している[3]。そして、自由度の拘束に関しては、二関節筋構造（一つの筋が二つの関節にまたがって接続されている構造）の脚ロボットでは、一関節筋のみの場合に比べ、運動のコーディネーションが容易であることが実験的に示されている。これらは、制御が身体構造と密接に結び付いていることを示している。すなわち、身体が環境との相互作用を通して、制御計算を担っているとも解釈できる。

その極端な例が、受動歩行[5]であろう。明示的な制御手法もアクチュエータもなしに、坂道で歩行を実現できる。これは、物理的身体のエネルギー消費（資源拘束や疲労）の観点からも重要である。

1.3.2.4 体表面

皮膚感覚は、その重要性の認識はありつつも、技術的な実現の限界から、人間型ロボットに、これまであまり採用されてこなかった。しかし、等身大ヒューマノイドの全身に柔軟かつ切り貼り可能な触覚センサを1800個以上実装することにより、ヒューマノイドの様々な身体部位と環境・対象物との接触を活用[6]した動作の実現を行った研究や、ロボットの柔らかい皮膚（シリコン製）の下に触覚センサとして約200個のポリフッ化ビニリデン（PolyVinylidene DiFluoride; PVDF）素子を実装したロボット

※2
入力されたエネルギーを物理的運動に変換する機械要素。

※3
「超多自由度の運動機構系に対して、どのように運動を構造化するか？」はニコライ・ベルンシュテイン(Nikolai Bernstein)氏が指摘した運動発達の基本問題である。超多自由度の空間内で、適切な運動を探索する際の空間は巨大なものとなり、単純な探索によって最適な運動を獲得するのに必要な計算量が膨大になってしまうことを指す。

※4
PETなどの繊維を編み込んだチューブに空気を入れると、風船のように膨らむが、一方で全長は収縮して短くなることを利用したアクチュエータ

※5
アクチュエータなどによるエネルギーの入力や、複雑な制御を行わなくても、きっかけを与えるだけで、脚の機構などにより、坂道などを歩行すること。

※6
ヒューマノイドの体のどこにいつ触れたのかについての情報を直接得ることにより、ヒューマノイドの動作安定性が増す。

のプラットフォームが開発されている。

また、全身ではないが、プラスチックの骨格にゴム手袋を装着し、PVDF素子とひずみゲージをシリコンと一緒に注入したバイオニックハンドを開発し、触覚センサによる指や掌の触覚と把持運動を利用して、数種の物体を識別する研究も行われている。センサをあらかじめ調整することはせず、創発により校正されることを目指している。人と比べるとセンサは圧倒的に少ないが、受容器の種類として類似の構造を取っており、人の把持スキルの学習発達研究への拡張が期待されている。

体表面の皮膚感覚は、体性感覚と密接に結び付き、自己の身体のイメージを獲得する上で非常に根源的かつ重要な感覚である[4]。高次脳機能がこのような基本的な知覚の上に構成されることを考えれば、知能発達の観点から、何らかの形で実装していることが望ましい。力学的な感覚受容器としての構造化に加え、痛みとしての感覚は、生物の場合、個体の生命維持に必須であるが、その社会的意味としての共感は、将来、人間と共生するロボットにも望まれる。その際、明示的にプログラムされた物理的インパクトへの応答ではなく、共感としての情動表現が可能であれば、より深いコミュニケーションが可能と考えられる。これは、以下のミラーニューロンシステムとも深く関連する。

1.3.3 知能の発達的設計

1.3.3.1 発達の多様性

発達の様相にはいくつかの視点がある[7]。一つの視点は、赤ちゃんの発達過程を外部から観測した場合、中央制御的ではなく、分散かつ創発的で漸次的過程とみなすことができる点である。

通常の人工的なシステムの場合は、明示的に設計されている構造と制御機構を基盤として上位の構造や機能が作り込まれるが、生物の場合は、不完全で効率の悪い構造と行動表現を基にして、上位の発達段階の構造が構築される点に大きな相違がある。また、乳幼児の生態学的な意味での拘束は、必ずしも不利な点ではなく、むしろ発達を促す。脳、身体、環境の間の共同作用若しくはパターン生成の固有の傾向は、各種「引き込み現象」[8]を引き起こし、更に、能動的探索により自己の身体表象、自由度の拘束などによる運動パターン生成など創発過程が見られる。

発達心理学においては、このような環境に対する能動的な探索と操作の結果として、知覚の範疇の獲得や概念形成が行われると考えられている。感覚やある種の知覚は運動とは無関係に処理されるが、知覚の範疇の獲得は感覚系と運動系の相互作用に依存する。これらの相互作用と創発過程を構成論的にモデル化するため、脳の微視的な構造や機能を参考として、各種の調整を行っている神経修飾物質[9]、神経可塑性[10]、強化学習などを計算モデルに組み込む研究も行われている。

もう一つの視点は、社会性がどのように獲得されるかという点である。巨視的なレベルでは、養育者を始めとする他者の関わりが、赤ちゃんの自律性、適応性、社会性を助長している。養育者による足場作り（scaffolding）は、認知的、社会的、技能的発達に重要な役割を果たす。また、乳幼児は養育者の反応に対する感受性期があり、養育者はこれに合わせて対応を調整する。人工システムにおいても、知能を発達させる観点から、何らかの形で社会的相互作用における養育者による足場作りに相当する環境設計が必要と考えられる。

※7
詳細は、文献[7]の第7章を参照されたい。

※8
引き込み現象とは、複数のものの運動が相互作用の中で特定のパターンに収束していく現象を指し、創発に係る一つのメカニズムであると考えられている。

※9
脳内の神経の振る舞いを制御する神経伝達物質のうち、時間的に持続的な効果を持つものの総称。

※10
神経系の構造が時間的に変化し得ること。

1.3.3.2 ミラーニューロンシステムと社会性発達基盤

社会的相互作用、特に養育者による足場作りにおいて、鍵となる役割を果たしていると考えられるのが「ミラーニューロン」と呼ばれる脳のニューロンである。ミラーニューロンは、ある行動を自分で行う場合と、他者が行う場合の双方に反応するニューロンであり、サル、人等で発見されたものである。人のミラーニューロンは、人の脳内で言語を扱う部位の近くに位置しており、言語能力に至る道筋での重要な役割を果たしていると推察されている。様々な研究から、ミラーニューロンに関わる多くの事柄が明らかになりつつある[5]。

ミラーニューロンシステムには、自己や他者の身体を認識するシステムがそれぞれ存在し、かつ一部を共有している。このことにより、ミラーニューロンシステムが、他者の動作プログラムを自身の脳内で再現すること、すなわち、他者の内部状態を自己の内部状態としてシミュレーションすることが可能となっている。これは、運動主体感、自他弁別、他者の行為認識とも関連すると考えられている。これらに基づき、以下のように考えられている[6]。

- ミラーニューロンはもともと、自他に関わらず、動作そのものを視覚的にコード化し、運動実行中に感覚フィードバックとして働いていたが、発達・進化の過程で、運動情報と統合され、現在のミラーニューロンを構成するようになった。
- 自己身体認知のステップとして、自分自身の運動指令の情報が脳の感覚情報を処理する領域に送られ（遠心性コピーと呼ばれる[11)]、感覚からのフィードバック（感覚フィードバック）と一致していれば運動の主体が自己であることを認識する（運動主体感の構成[12)]）。
- 遠心性コピーと感覚フィードバックが一致しない場合には、その運動主体感が構成されず、感覚信号は他者の身体による行為の結果と認知される。
- ミラーニューロンは、他者の動作認識とともに、自己の身体や他者の身体の認識にも関与している。

更に、ミラーニューロンシステムは、模倣、共同注意、心の理論、共感、コミュニケーションなどの他者の意識や心理の状態を把握する機能と関連すると考えられている。これらは人に特有の自己と他者の共通性と差異に基づいた自己や他者への気づきの駆動、社会的な行動の学習・発達に寄与しているとみなされている。

実際、サルは模倣しないと言われている。また、サルのミラーニューロンシステムの場合、対象が明示された他動詞的な動作にしか反応しないのに対し、ヒトの場合、自動詞的な動作、つまり目的を持たない行動に対しても反応するミラーニューロンシステムが存在する[5]。

サルの場合、人に比べて、個体の生存のための圧力が大きいので、ゴール指向の運動が個別に確立して、いち早く駆動することが可能である。それに対して、人の場合、養育者の庇護を受けるので、その圧力が小さく、ゴール指向のみならず、目的を持たない要素運動的なものにも反応することで、学習による構造化や組織化による汎用性が高まる余裕があり、結果として、より社会的な行動や認知能力へ拡張されたと考えられる。

※11
運動制御においては運動の指令が運動野に送られるだけでなく、その信号のコピーが感覚野に送られていると考えられている。このコピーは、中枢神経から末梢系に送られるので、遠心性コピーと呼ばれている。

※12
文献[6]では、実際、頭頂葉のニューロンが遠心性コピーと感覚フィードバックの情報の統合に関わることを発見している。

1.3.3.3 発達的な知能の設計論

身体性の意味と役割、ロボットの学習を概観してきたが、これらの考え方や既存科学の知識や知見を鑑み、人工システムの知能の発達的な設計論として、より系統的な枠組みが必要と考えられる。そこには、古典的課題である「氏と育ち」の課題と関係が、設計論を通じて浮き彫りにされる。

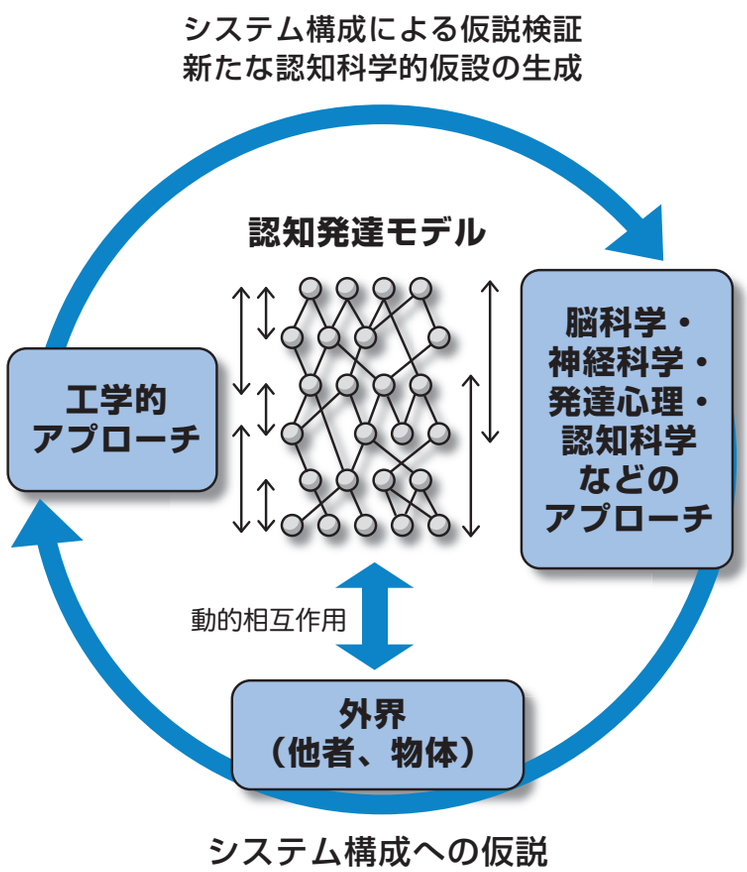
氏と育ちが対立概念ではなく、育ちを通じて氏が形成される（Nature via Nurture）と主張するマット・リドレー（Matt Ridley）氏に習えば、人工システムも同様に、事前の埋め込みと、それに基づく学習・発達系の設計論が必要である。そのような思想背景から、以降で紹介する認知発達ロボティクスの考え方が生まれ、構成的発達科学へ拡張される。これが、知能の発達的設計の基盤である。

1.3.4 認知発達ロボティクス

1.3.4.1 認知発達ロボティクスの基本的な考え方

認知発達ロボティクスとは、従来、設計者が明示的にロボットの行動を規定してきたことに対し、環境との相互作用から、ロボットが自ら行動を学習し、それらを発達させていく過程に内包される抽象化、概念の獲得を実現するためのロボット設計論である。

認知発達ロボティクスの焦点は、自律的な主体（エージェント）が環境との相互作用を通して、世界をどのように表現し、行動を獲得していくかといった、ロボットの認知発達過程にある。特に、環境因子としてほかのエージェントの行動が自分の行動をどのように規定していくかという過程の中に、ロボットが「自我」を見出していく道筋が解釈できるのではないかという期待がある。このように環境との相互作用をベースとして、その時間的発展に焦点をあて、脳を含む自己身体や環境の設計問題を扱う研究分野が認知発達ロボティクスである。

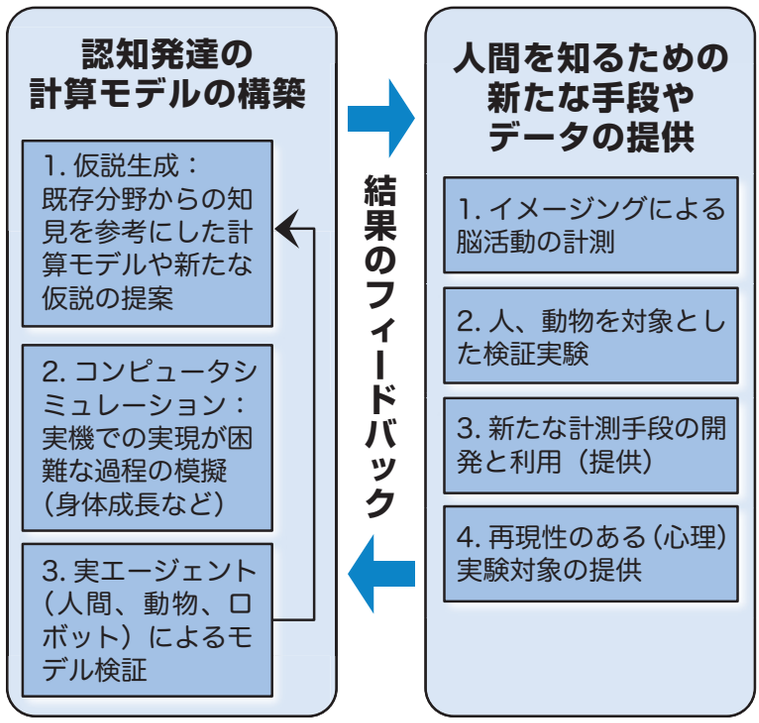


■図13　認知発達ロボティクスの概念

認知発達ロボティクスの基本的な考え方は、問題自体に対する理解の過程を、ロボット自身が環境との相互作用を通じて経験することにより、様々な状況に対応可能なメカニズムを構成論的アプローチによって構築することである。特に、知的行動を人間のレベルまで求めるのであれば、人間以外の動物にも可能な連合学習[13]のレベルから、人間特有の記号の生成と学習、すなわち言語獲得に至る過程（言語創発）が（1.2節参照）、ロボットの内部構造と外部環境の多様かつ制約的相互作用の中に見出さなければならない。

※13
2種類の刺激の組合せを学習すること。

従来のロボティクスでは、人間と共生するロボットのコミュニケーション技術として、トップダウン的に言語構造を与えたがために、言語創発過程が内包されていない。それゆえ、表層的な言語コミュニケーションに留まり、限られたコンテキストでの定型的な応答しかできない。その一方で、認知発達ロボティクスでは、言語創発に至る過程そのものを人工的に構成することで、人の認知発達過程の理解とともに新たなロボット設計論を目指す。



■図14　認知発達ロボティクスやそれに関連するアプローチ

このような人の認知に関する研究は、従来、認知科学、神経科学、心理学などの分野で扱われてきた。そこでは、説明原理による理解を目指しており、認知発達ロボティクスが志向する設計原理に基づくものではない。しかしながら、人間理解という共通基盤を基に、工学的アプローチからは、「システム構成による仮説検証や新たな認知科学的仮説の生成」が、認知科学、神経科学、心理学などの分野に提案され、逆に、これらの分野から、「システム構成への仮説」が工学的アプローチに提案され、相互フィードバックによる認知発達モデルの構成と検証が可能である。それが認知発達ロボティクスの一つの理想形である（**図13**）。

1.3.4.2 認知発達ロボティクスの設計論

認知発達するロボットの設計論は、環境、身体、タスクが一体となって構成されなければならない。ここでは二つに分けて説明する。一つは、身体を通じて行動するための環境表現を構築していくロボットの内部の情報処理の構造をどのように設計するか、もう一つは、そのように設計されたロボットが上手に学習や発達できるような環境、特に教示者を始めとする他者の行動をどのように設計するか、である。両者が密に結合することで、相互の役割である学習・発達が可能である。

重要なポイントは、獲得すべき行動をロボットの脳に直接書き込むのではなく、他者を含む環境を介して（社会性）、ロボット自身が自らの身体を通じて（身体性）情報を取得し解釈していく能力（適応性）と、自らその過程を駆動できることである（自律性）。

認知発達ロボティクスやそれに関連するアプローチ[14]は、まだその事例が少ないが、その方向性は、主に二つに分かれる。一つは、機構の仮説を立て、コンピュータシミュレーションや実際のロボットを使って、実験し、仮説検証と仮説の修正を繰り返すことである。もう一つは、脳活動の計測や人、動物を対象とした検証実験等により人間を知るための実データを提供することである。これらは互いに関連し、相互フィードバックし得る（**図14**）。

※14
「JST ERATO Asada Project」科学技術振興機構ウェブサイト
<http://www.jst.go.jp/erato/asada/>

1.3.4.3 記号創発ロボティクス

人間の認知発達を考えた際に、言語獲得は特に重要である。それは、言語を人間がコミュニケーションに用いるからという理由だけではなく、思考や推論など様々な高次の認知過程に言語が用いられると考えられているためである。それゆえに、言語は人間と他の動物を分かつ重要な要素であるとみなされてきた。数理論理学を基礎として生まれた初期のAIにおいては、記号処理、つまり記号で表現された述語論理式や変数を操作することが、知能の本質であるとみなされてきた。

だが、その後、記号操作だけでは実世界の不確実性やダイナミックな環境に対応できないことが指摘された。また、記号の意味をどうやって現実世界に関係付けるのかという問題はシンボルグラウンディング（記号接地）問題と呼ばれ、AIの基本問題の一つに位置付けられる。

しかし、言語の体系に代表される記号システムそれ自身が、身体に基づく経験や社会的相互作用の下で形成されるとするならば[15]、人間がトップダウンにAIに与えた記号システムを絶対的に真な記号システムとして取り扱うことには無理が生じるし、記号システムの適応性を欠いてしまう可能性もある。記号システムはそれ自体が各認知主体の環境適応と、他者とのコミュニケーションを経て、創発的に形成されたと考えるのが妥当であろう。

このような視点から、AIの設計を記号システムありきで始めずに、記号システムが環境適応を通して形成されるプロセスを機械学習により表現し、言語獲得や、社会における言語の形成を、ロボットを用いて構成論的に理解しようとするのが記号創発ロボティクスという研究分野である。

記号創発ロボティクスは、人間のコミュニケーションを支える社会的システムである記号が創発されるシステム（記号創発システム）への構成論的アプローチであると位置付けられている[8]。AIにおいては、身体に近い感覚運動系の低次の認知と言語や論理といった高次の認知をいかに橋渡しするかという問題は、長きに渡って存在してきた。だが、記号創発ロボティクスは、それを感覚運動系からのボトムアップな学習によって、高次の認知まで説明しようとするアプローチであるといえる。

認知発達の一側面である言語獲得を中心に、ロボットを用いて認知発達の過程に構成的に迫るという側面において、記号創発ロボティクスは認知発達ロボティクスの一つの支流である。ただし、人間社会

■図15　神経ダイナミクスから社会的相互作用へ至る過程の理解と構築による構成的発達科学プロジェクトの概要

※15
各人が個別に記号システムを持っている状態では、他者とコミュニケーションを取ることができない。コミュニケーションを行うためには、社会的に共有されている記号の体系が必要となる。ここではそのような体系を指して記号システムと呼んでいる。文献[8]参照。

における記号システムの記号創発のメカニズムを背景とした理解に重きを置きながら、方法論としてはディープラーニングなどの機械学習を駆使して、認知発達における言語獲得の過程を構成的に研究する点に特徴がある。

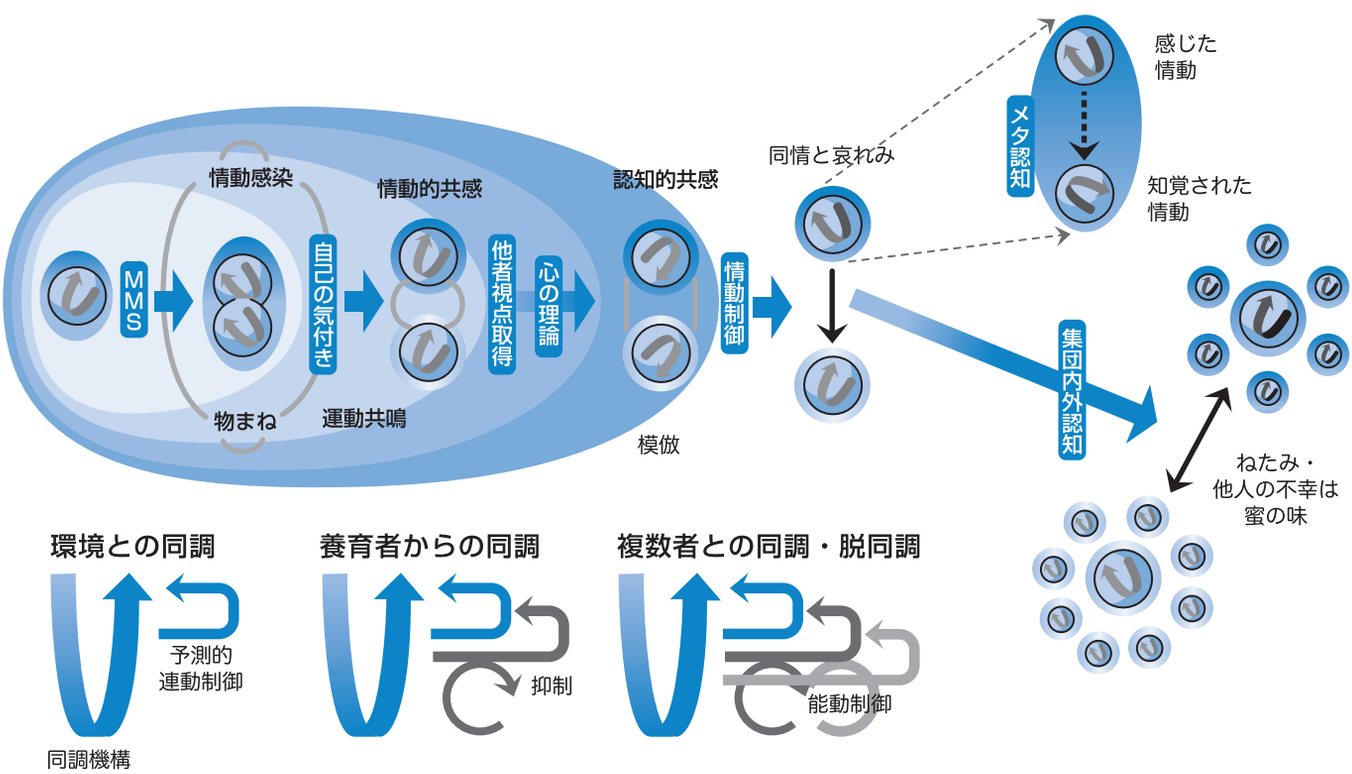
1.3.5 構成的発達科学

知能設計の研究には、機械学習やロボティクスだけでなく、神経科学、発達心理学、認知科学など周辺の幅広い諸科学が深く関係する。これらの諸科学を発達の研究のために総合する学問が構成的発達科学であり、記号創発ロボティクスやそれを含む認知発達ロボティクスの考え方をより発展させたものである。

構成的発達科学の主な研究例として、神経ダイナミクスから共感や自他認知の発達などの社会的相互作用へ至る過程の理解と構築を目指した研究と、胎児の発達原理に基づいて発達障害をシステム的に理解しようとする研究[9][10]を、以下に紹介する。

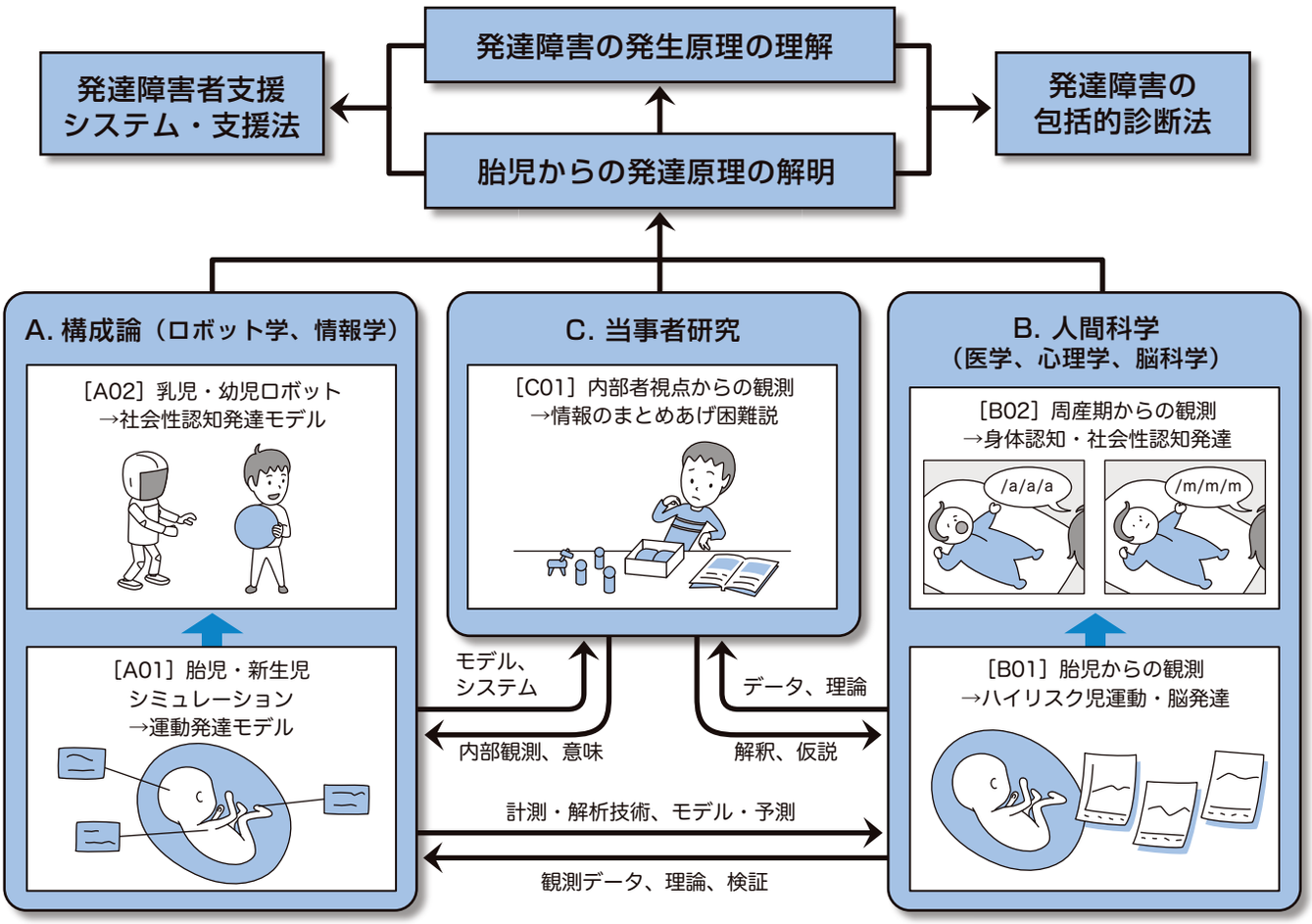
前者の研究では、ロボットプラットフォームを用いた自他認知に関わる心理・行動実験を行っている際の、被験者の脳機能イメージングを行うことにより、計算モデルを構築し、自他認知の発達原理を解明することを目指している。研究全体の簡単な概要を**図15**に示す。

自己認知の発達過程は、人工システムが何をもって、自己の概念を獲得したかの判別が困難である。そこで、共感の発達になぞらえて、人工共感を設計することを想定した模式図に自他認知の発達が組み込まれ、具体的な心的機能が目標として与えられている。**図16**にその概念図を示す。具体的には、脳



■図16　共感発達モデルとしての自他認知過程[16]

※16
文献[7]より作成。



■図17　構成論的発達科学ー胎児からの発達原理の解明に基づく発達障害のシステム的理解ープロジェクトの概要

機能イメージングにより、親子ペアが相互に顔を見つめているときと、無関係なビデオを鑑賞しているときの差違が、自閉症軽度と重度の場合で異なることが見出されている。

後者のプロジェクト（**図17**）では、発達障害者の当事者研究を実施している点が特徴的である[9][10]。当事者研究とは、発達障害者が自らの感覚や経験を観測して体系的に記述することで、モデルの検証や意味付けを行う研究の手法である。例えば、自分が「おなかが空いている」ことを無意識的に感知できない発達障害者がいる。彼らは、複数の下位レベルの状態の集合として意識的に情報をかき集め、自分がどのような状態であるかを推定することにより「おなかが空いている」ことを知覚している。この過程は、先に述べた無意識的過程に相当すると考えられるため、彼らの記述からこの過程をモデル化する際の示唆を得ることができる。

1.3.6 ロボット学習としてのディープラーニング

ディープラーニングは、階層的な特徴量[17]の学習により、“ほとんど同様の枠組みで多様なモダリティ[18]を扱える”という特長をもち、ロボット学習への応用が盛んに行われている。実世界で行動するロボットシステムは通常、カメラ、マイクロフォン、接触センサ、アクチュエータなどを備えたマルチモー

※17
特徴量とは、問題の解決に必要な本質的な変数であったり、特定の概念を特徴づける変数のことを指す。複数の種類のセンサ信号やアクチュエータを備えるロボットの場合、それぞれの入出力の特徴量を更に階層的に組み合わせることで、統合的な認識が可能となる。

※18
視覚、聴覚、触覚等の感覚の種類のこと。

ダルシステムであり、ディープラーニングが有効に利用できる。

1.3.6.1 ロボットビジョン

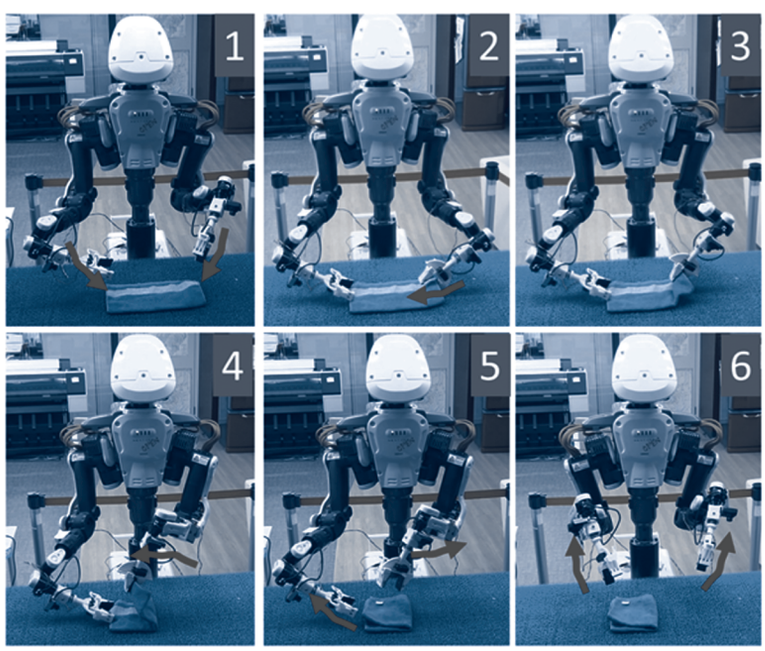
ロボットビジョンはディープラーニングが最初に導入された応用分野である。例えば、ジェヨン・スン（Jaeyong Sung）氏らは、ディープラーニングにより対象物の距離画像から把持ベクトル（ロボットハンドの位置と方向）を出力させる手法を提案している。ジョセフ・レドモン（Joseph Redmon）氏とアネリア・アンジェローヴァ（Anelia Angelova）氏らは色（RGB）と深さ（Depth）データから構成される三次元RGB-D画像から、ディープラーニングモデルの一種である畳み込みニューラルネットワーク（CNN）を利用し、把持ベクトルの予測を行っている。

これらの研究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扱っている点で、純粋なビジョン研究であると言える。実際の動作自身は、把持ベクトルの情報を受けて、逆運動学[19]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る。

1.3.6.2 動作学習（End to End Learning）

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のかが、当然考慮されなければならない。つまり、学習対象には、物体の画像のみならず、動作を生成する身体の構造が含まれる必要がある。

ディープラーニングの重要な方法論の一つに、入力から出力までを一つのネットワークとして表現し、全体を学習してしまう「End to End Learning」がある（1.2.8項参照）。従来は、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が必要であったが、ディープラーニングにより、全体のネットワークが一気に



■図18　Nextageによる柔軟物の折り畳み動作の学習

※19
ロボットや3Dモデルの関節を制御する方法であり、逆運動学とは指先やつま先の位置から関節の角度が決まることをいう。

※20
Willow Garage（米国）が開発、販売している双腕ロボット。ハードウェアとソフトウェアがオープンプラットフォームで開発されている。 Willow Garage Website <http://www.willowgarage.com/pages/pr2/overview>

学習できるようになった。ロボットの動作学習の場合には、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）を直接得るという発想となる。この方法論を適用することで、実際に実行可能な動作の学習が可能となる。

セルゲイ・レヴィン（Sergey Levine）氏らは、ロボットPR2[20]にディープラーニングによる行動探索を行わせ、現時刻の一枚の視野画像の入力から次時刻のロボットの複数関節を直接CNNで出力させ、一連の動作を実現する手法を提案している。複数の動作について、物体の位置変化などがあっても、安定的に動作が行えることを示した。

また、同様の手法を14台のロボットアームに導入し、80万回のピッキング動作の学習によって多様な一般物体のハンドリングを実現している。またピンチェ・ヤン（Pin-Che Yang）氏は、「Programming by Demonstration」（動作教示からの動作生成メカニズムの獲得）の視点から、ディープラーニングにより、「Nextage」[21]を用いて画像時系列からの未学習のタオル折り畳み動作生成、及び将来の視野画像の予測などを行っている（**図18**）。またリカレントニューラルネットワーク（RNN）を利用することで、動作の安定化を実現している。

1.3.6.3 言語学習

画像や映像からのキャプション生成など、ディープラーニングによる言語とほかのモダリティとの統合手法はロボットにも応用可能である。この視点から、ディープラーニングモデルによるロボットの動作に関する言語の利用や認識に関する試みもいくつか報告されている。

例えば、イェヅォウ・ヤン（Yezhou Yang）氏らは、人間の調理映像にその状況を説明する言語情報を加えてCNNに与えることで作業シーンの分節化を助け、オブジェクト48種類と6種類の把持タイプを識別し、ロボットの動作に適用している。山田竜朗氏らは、ディープラーニングによる自然言語処理に利用されるRNNの「Sequence to Sequence Learning」を応用し、小型ロボット「NAO」の物体操作タスクにおいて、状況依存性（多義性）を持つ未学習の文章指令から運動への変換を実現している。

1.3.7 歴史的経緯と国内外の研究動向

1.3.7.1 歴史的経緯

認知発達ロボティクスの提唱以前のAIやロボティクスでは、モデルベーストや記号表象に基づく手法が主流であった。これに対し、1984年にヴァレンティノ・ブライテンブルク（Valentino Braitenberg）氏が、簡単な構造を持つ光センサとモータを持つ移動ロボットが、複雑な行動を生成し得ることを示した。また、1986年当時マサチューセッツ工科大学（MIT）にいたロドニー・ブルックス（Rodney Brooks）氏は、知能内部の精巧な推論プロセスを作り込む必要はなく、環境との相互作用により知能が実現されるとする「表象なき知能」という考え方を提唱し、実際に昆虫を模したロボットで障害物回避や不整地歩行が可能であることを示した[11]。

その後ブルックス氏は「Attila」、「Hannibal」など、1990年代にかけて様々な昆虫型のロボットを開発している。元チューリッヒ大学（スイス）のロルフ・ファイファー（Rolf Pfeifer）氏は1987年にAI研究所を設立し、身体性に重きをおいた研究を開始している。

また、1990年代にはロボカップの開催（1997年に第1回大会が開催）、経済産業省の「人間協調・共存ロボットシステムの研究開発」プロジェクトから開始されたHRPシリーズの開発（1998年～）等、

※21
川田工業が開発した人と共存する環境で作業することを念頭に置いた汎用の人型ロボット。

実際のロボットを用いて人間とロボットとの協調動作やロボット同士の協調動作を目指した研究が実施された。これらを背景に、先にも述べた認知発達における身体性や社会性の重要性が認識され始め、「身体性認知科学」や「認知ロボティクス」「認知発達ロボティクス」が勃興し、知の理解と創造が不可分であるとの認識が高まった。

1.3.7.2 国内の動向

国内では、2005年秋から科学技術振興機構（JST）の戦略的創造研究推進事業（ERATO）「浅田共創知能システムプロジェクト」が発足した。約5年半に渡り、ヒューマノイドロボットの新たな設計・製作・作動と認知科学や脳科学の手法を用いた構成モデルの検証による、科学と技術の融合した新領域「共創知能システム」を構築することを目標に、研究が展開された。このプロジェクトでは、人工筋肉などの柔軟素材を用いた身体構造と環境と動的結合による運動の創発、身体的行動から対人コミュニケーションまでを発達的につなぐ認知モデルによる認知発達の構成的理解、アンドロイドやマルチロボットシステムを用いたコミュニケーションの理解と実現、脳機能画像計測や動物実験による構成モデルの検証などを中心に行われた。

上記のプロジェクトの動きと並行して、KAIST（韓国）の谷淳氏らがRNNを用い、複雑系科学の分野で、身体性認知科学としてのロボットの神経ダイナミクスの研究を究めている。早稲田大学の尾形哲也氏は、RNNやディープラーニングを駆使したロボットの行動学習に焦点を置いて2014年頃から研究しており、ロボットにマルチモーダル学習をさせ、指示によって行動を切り替えさせることに成功している。立命館大学の谷口忠大氏は、記号創発ロボティクス（1.3.4項参照）を提唱し、認知発達ロボティクスにおける言語創発の課題に焦点をおいた研究を実施している。谷口氏を含め、玉川大学の岡田浩之氏、早稲田大学の尾形氏がメンバーになり、電気通信大学の長井隆行氏が率いる「記号創発ロボティクスによる人間機械コラボレーション基盤創成」プロジェクト[22]が2015年から始まっている。

また、身体性をテーマとした研究は現在も活発に実施されている。体表面（1.3.2項参照）の高度化を目指し、ソフトロボティクスにおける人とのインタラクションを科学としてとらえることを狙ったJSTさきがけプロジェクト「触れ合いデータを収集する子供アンドロイド高機能化」[23]（2016年～）、ミラーニューロンシステム（1.3.3項参照）の予測符合化[24]を基にしたモデル化と発達障害者支援を目標としたJST CRESTプロジェクト「認知ミラーリング：認知過程の自己理解と社会的共有による発達障害者支援」[25]（2016年～）や、アンドロイドの社会的行動の工学的実現を目指したJST ERATO「共生ヒューマンロボットインタラクションプロジェクト」[26]（2016年～）などが継続的に推進されている。

1.3.7.3 海外の動向

海外では、1987年にAI研究所を設立したファイファー氏が、身体性に重きをおいた身体性認知科学の多くの研究や思想をまとめて、書籍として著している。柔軟な身体が有する環境との相互作用の豊かさが、いかにして知の創造や理解に繋がるかを説いている。これら、一連の研究は認知発達ロボティク

※22
「記号創発ロボティクスによる人間機械コラボレーション基盤創成」プロジェクト. 科学技術振興機構ウェブサイト <http://www.jst.go.jp/kisoken/crest/project/1111083/15656632.html>

※23
「【石原　尚】触れ合いデータを収集する子供アンドロイド高機能化」科学技術振興機構ウェブサイト <http://www.jst.go.jp/kisoken/presto/project/1112079/1112079_02.html>

※24
脳内で未来の感覚情報の予測が行われていること。

※25
科学技術振興機構　戦略的創造研究推進事業「認知ミラーリング：認知過程の自己理解と社会的共有による発達障害者支援」プロジェクト. 科学技術振興機構ウェブサイト <http://cognitive-mirroring.org>

※26
科学技術振興機構ERATO「共生ヒューマンロボットインタラクションプロジェクト」<http://www.jst.go.jp/erato/ishiguro>

スにおけるロボットプラットフォームの重要性、更には、認知過程への本質的な貢献の役割も含めて、現在では、ソフトロボティクスの思想基盤となっている。

イタリアでは、ジェノバ大学のジュリオ・サンディーニ（Giulio Sandini）氏が、コンピュータビジョンの研究で、当初から生体視覚に着目し、物体の把持や操りなど、行動系と密着したロボットビジョンの研究を行ってきた。そして、発達の概念を取り入れた認知発達ロボティクスの考え方を、2003年にファイファー氏らと一緒に著している。彼らは、「iCub」と呼ぶ幼児ロボットプラットフォームを開発し、認知発達研究のプラットフォームとして、世界に送り出している。サンディーニ氏は、その後、イタリア技術研究所（IIT）のロボット領域の主要メンバーとなり、人間の行動系の研究を中心に多くの研究成果を挙げている[27]。

英国では、プリマス大学のアンジェロ・ケンジェロシ（Angelo Cangelosi）氏が、iCubを用いた言語発達研究を推し進めている。これは、ヨーロッパのITALK（Integration and Transfer of Action and Language Knowledge in Robots）プロジェクト（2008年3月～2012年2月）の一部であり、IITや後述のビーレフェルト大学（ドイツ）など、ヨーロッパの主要な研究機関が含まれていた[28]。ケンジェロシ氏は、南イリノイ大学の心理学者マシュー・シュレシンガー（Matthew Schlesinger）氏とともに、『Developmental Robotics － From Babies to Robots －』と題する書籍を著しており、その中にITALKプロジェクトの成果も含まれている。また、多くのプロジェクトに参画している[29]。

ドイツでは、ビーレフェルト大学のヘルゲ・リッター（Helge Ritter）氏が、HRI（Human Robot Interaction）を中心とした、工学実利的な側面に重きをおいた多くのプロジェクトを1990年代から現在まで長年に渡って遂行しており、EUにおけるCOEプログラム（Center Of Excellence Program）を獲得し続けている。また、CITEC（the Cluster of Excellence Center in Cognitive Interactive Technology）と呼ばれる研究センターを2007年に設立し、多くの研究者を抱えて活動している[30]。

最近の大きなプロジェクトの一つは、「The Cognitive Service Robotics Apartment as Ambient Host」と呼ばれているもので、2013年10月からの4年プロジェクトで家庭内環境での認知ロボットの活動の実現を目指している。フランクフルト大学のFIAS（Frankfurt Institute of Advanced Studies）のヨッフェン・トリーシュ（Jochen Triesch）氏は、もともとコンピュータビジョンの研究を行ってきたが、神経科学をベースに認知過程のモデル化を試みている。すなわち、脳のネットワークの創発により知的な感覚と行動が生成される機序を明らかにしようとしている[31]。

フランスでは、国立情報学自動制御研究所（INRIA）のピエール・イヴ・ウーディユ（Pierre-Yves Oudeyer）氏率いる研究グループでは、内発的動機付けを情報論の立場から明らかにしようとする発達ロボティクスの研究を行っている[32]。セルジーポントワーズ大学（フランス）のフィリップ・ゴシエ（Philippe Gaussier）氏は、脳神経系のモデル化を通じた認知発達過程の解明を目指している。脳の各部の機能とその関係を計算モデルとして具現化し、ロボットによる検証を通じて、新たな理解と洞察を

※27
“Giulio Sandini.” Italian Institute of Technology Website <https://www.iit.it/people/giulio-sandini>
※28
Integration and Transfer of Action and Language Knowledge in Robots Website <http://www.italkproject.org>
※29
“Angelo Cangelosi: Professor of Artificial Intelligence and Cognition.” University of Plymouth Website <http://www.tech.plym.ac.uk/soc/staff/angelo/>
※30
The Cluster of Excellence Center in Cognitive Interactive Technology.Website <https://www.cit-ec.de/en/citec>
※31
“Research Group of Jochen Triesch.” Frankfurt Institute of Advanced Studies Website <http://fias.uni-frankfurt.de/neuro/triesch/>
※32
Pierre-Yves Oudeyer Website <http://www.pyoudeyer.com>
※33
Equipes Traitement de l'Information et Systèmes Website <http://perso-etis.ensea.fr/gaussier/>

得ることを狙っている[33]。

米国では、1991年にブルックス氏が行動規範型ロボットのアーキテクチャを提唱後、一連の研究が行われてきたが、認知発達ロボティクスに強く関連する研究グループの形成には至らなかった。その後、ブルックス氏は、お掃除ロボット「ルンバ」で有名なiRobotを設立・創業後、Rethink Roboticsを設立し、新たな産業用ロボット「Baxter」の開発・販売に従事している。

認知発達ロボティクスのアプローチの一つは、ロボットを道具として人間研究に利用することである。シアトルにあるワシントン大学で、ILABS（Institute for Learning and Brain Sciences）を率いる心理学者夫妻のアンドリュー・N・メルトゾフ（Andrew N. Meltzoff）氏とパトリシア・K・クール（Patricia K. Kuhl）氏は、赤ちゃんの行動実験及び脳磁計（Magnetoencephalography; MEG）による計測を通じて、発達研究を行っているが、ロボットやモデル研究との連携にも強い関心を示している。そして、フランスのゴシエ氏との共同研究や、大阪大学の浅田稔氏からロボットを譲り受けて、ロボットの学習や社会性に関する研究を行っている。インディアナ大学（米国）は、発達心理学で著明なリンダ・スミス（Linda Smith）氏が、計算モデルに関心を持っており、ユ・チェン（Chen Yu）氏と連携し、赤ちゃんの発達モデル化を試みている。

脳神経系の構造をベースにロボットの感覚行動のマッピングを対象とする研究は、「Neurorobotics」と呼ばれ、ノーベル賞受賞者のジェラルド・モーリス・エデルマン（Gerald M. Edelman）氏が、1990年代初期から行ってきた。エデルマン氏は、脳のネットワーク構造を明らかにしようとするコネクトーム[34]のオラフ・スポーンズ（Olaf Sporns）氏や意識の研究で著明なジュリオ・トノーニ（Giulio Tononi）氏と共同研究を実施している。Neuroroboticsとしての後継者は、ジェフリー・L・クリッチマー（Jeffrey L. Krichmar）氏である。エデルマン氏のグループ出身のユージン・M・イジケヴィッチ（Eugene M. Izhikevich）氏（Brain corp、米国）は、大規模な脳活動シミュレーション研究を行っている。これら一連の研究は、脳研究の派生と見なせるが、既存の脳研究のアプローチの限界を打破する上で、人工物設計を通じたアプローチは、認知発達ロボティクスの理念と通じる。

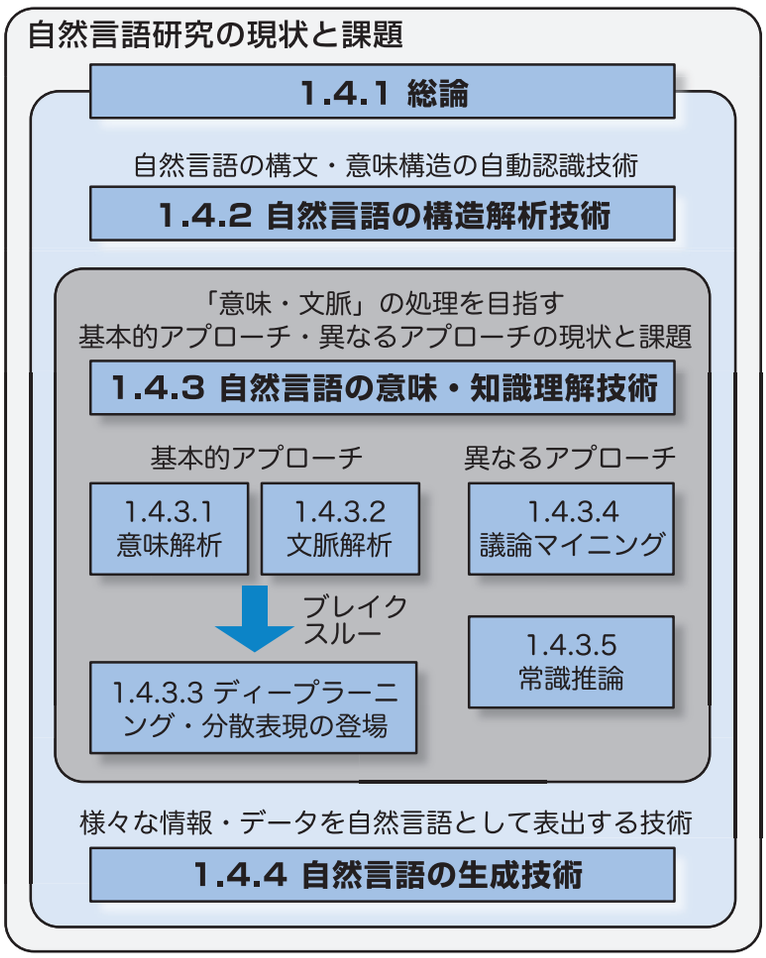
日本以外のアジア地域では、残念ながら、認知発達ロボティクスの研究はあまり行われていない。先に述べた谷氏らのグループが韓国のKAISTで行っている程度である。

参考文献
[1] 浅田稔・國吉康夫『ロボットインテリジェンス』岩波書店.
[2] Donald A. Neumann（嶋田智明・平田総一郎監訳）『筋骨格系のキネシオロジー』医歯薬出版.
[3] 柿谷慧ほか『筋骨格ロボットを用いた跳躍運動の学習』「ロボティクスシンポジア予稿集」vol.14, pp.380-385.
[4] サンドラブレイクスリー・マシューブレイクスリー（小松淳子訳）『脳の中の身体地図―ボディ・マップのおかげで、たいていのことがうまくいくわけ』インターシフト.
[5] ジャコモ・リゾラッティほか『ミラーニューロン』紀伊国屋書店.
[6] 村田哲ほか「脳の中にある身体」『ソーシャルブレインズ　自己と他者を認知する脳』東京大学出版会，pp.79-108.
[7] Minoru Asada, “Towards articial empathy,” International Journal of Social Robotics, vol.7, pp.19-33.
[8] 谷口忠大『記号創発ロボティクス　知能のメカニズム入門』講談社，2014.
[9] 綾屋紗月・熊谷晋一郎『発達障害当事者研究－ゆっくりていねいにつながりたい』医学書院.
[10] 綾屋紗月・熊谷晋一郎『つながりの作法-同じでもなく　違うでもなく』NHK出版.
[11] R.Pfeifer・J.Bongard（細田耕・石黒章夫訳）『知能の原理　身体性に基づく構成論的アプローチ』共立出版.

※34
ニューロンの間の接続状態を表した神経回路の地図のこと。

1.4 自然言語を中心とする記号処理

1.4.1 総論



■図19　本節の構成

人間は情報を抽象化・記号化し、更にそれを組み合わせて複雑かつ複合的な情報を表現・理解する能力を持っている。情報の抽象化はおそらく人間以外の生物も行っているだろうが、情報を記号化して組み合わせることで自由自在に加工・利用する能力、つまり自然言語をあやつる能力は、おそらく人間に限られたものであろう。人間であればだれでも当たり前に行っているこの能力をコンピュータ上で再現することが、記号・言語を対象とする人工知能（AI）研究の最終目標である。

自然言語に関する研究では、離散的構造と統計的性質、入力データと背景知識、パターン認識と論理推論など、異なる性質を統合的にモデル化することが求められる。特に最近は、AIの他分野と同様にディープラーニングを利用する研究が目立ち、これまではSVM（Support Vector Machine）やCRF（Conditional Random Field）を用いるのがスタンダードであったものが、フィードフォワードニューラルネットワーク（FFNN）やリカレントニューラルネットワーク（RNN）に置き換わりつつある。

ただし、これは前述の様々な性質の一面について強力な解を提供するものの、自然言語のモデル化の全てを解決するものではない。実際、自然言語処理においてディープラーニングを利用することによる成果は、今のところ様々である。機械翻訳や画像説明文生成のように大幅な性能向上が達成されているもの、構文解析や意味解析のようにインクリメンタルな精度向上は見られるものの基本的な手法はあまり変わらないもの、文脈解析や常識推論など現在のアプローチでは実用的な精度は見込めないもの、などがある。

特に、自然言語の意味・文脈理解や、対話システムにおける対話制御など、タスク設定やモデル化が模索されている段階の研究分野では、課題の多くは、現在も未解決である。

以下では、自然言語に関する研究のうち、自然言語文の内部構造（構文構造、意味構造）を認識する構造解析技術（1.4.2項参照）、自然言語がエンコードしている情報を理解・活用する意味・文脈解析技術（1.4.3項参照）、情報をエンコードして自然言語文として表出する自然言語生成技術（1.4.4項参照）、の三つについて、現在の技術動向と今後の展望について述べる。

1.4.2 自然言語の構造解析技術

1.4.2.1 構文解析

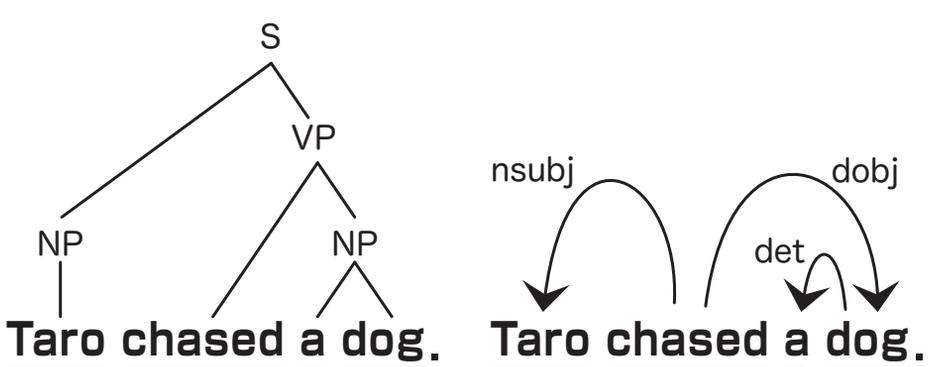
自然言語は、単語を並べることで複雑な情報を表現することができる。しかし、単語の並びが情報をエンコードする方法は自明でない。例えば、「Taro chased a dog.」と「A dog chased Taro.」という二つの文では、使われている単語は同じであるが、だれがだれをchaseするのか、という情報は反対になっている。英語では、主語、目的語といった役割が語順によって決められるため、語順を入れ替えるとこのようなことが起きるのだ。

一方、日本語では「が」や「を」といった助詞が主語・目的語を決めるため、「太郎が犬を追いかける」と「犬を太郎が追いかける」は同じ意味を表す。したがって、自然言語データを単なる文字列としてではなく、自然言語を用いて人間が理解している「情報」、すなわち「意味」をコンピュータで扱うためには、自然言語文の背後にある構造を明らかにする必要がある。

自然言語の文の構造を解析する技術は「構文解析」（syntactic parsing）とよばれ、古くから多くの研究がなされてきた。自然言語は再帰的な構造（「花子が飼っている犬が太郎を追いかけた」といった文のように、一つの文の中に更に別の文が入る、という構造）を持つため、文の構造を表現するためには木構造が用いられる。特に、自然言語の構文解析では、**図20**に示す句構造（phrase structure）と依存構造（dependency structure、係り受け構造ともいう）が広く用いられる。句構造は、単語列の構文的なまとまり（S：文、VP：動詞句、NP：名詞句など）を木構造で表す。依存構造は、単語の間の文法関係（nsubj：名詞主語、dobj：直接目的語、det：限定詞など）を木構造で表す。

構文解析では1990年代半ばから、統計的機械学習を用いる手法が主流である。機械学習の観点からすると、構文解析とは系列データ（単語列）に対して木構造（句構造あるいは依存構造）を推定する問題である。これには、木構造の「良さ」（正解の木構造との近さ）を定量化する問題と、良い木構造を効率的に探索する問題が含まれる。

前者については、確率文脈自由文法やそれを拡張したものや、SVM、最大エントロピー法、単層パーセプトロン、CRFといった線形分類器がよく使われた。最近では、FFNNやRNNなどのディープラー



■図20 「Taro chased a dog」に対する句構造（左）と依存構造（右）

ニングを適用することで更に精度が向上している。後者については、CYK（Cocke Younger Kasami）法やチャート法といった動的計画法、遷移型解析アルゴリズムなど木構造を系列ラベル付け問題に帰着する手法、最良優先探索やA*探索といったヒューリスティック探索など、AIの基本技術を応用したものが多い。

構文解析の最先端の研究では、機械学習手法と木構造の探索手法のより良い組合せを探求することで、少しずつではあるが解析精度が着実に向上してきている。その結果、英語や日本語といった言語においては、2000年代には90%以上の解析精度が達成された。

現在、構文解析の研究は、構文木の正解データ（「ツリーバンク」という）を学習データとした教師あり学習が主流である。ツリーバンクの開発には多大なコストと時間が必要であり、大きなツリーバンクが利用できる言語は限られている。上述のように英語や日本語において高精度な構文解析が実現されているのは、「Penn Treebank」や「京都大学テキストコーパス」[1][1]といった大規模ツリーバンクに負うところが大きい。

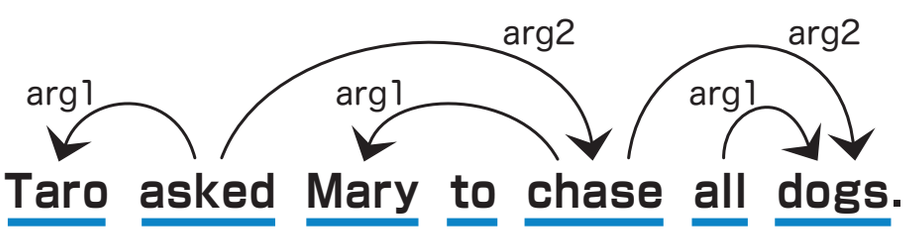
最近では、多数の言語について同じ基準でツリーバンクを開発する「Universal Dependencies」プロジェクトが注目されており、これまでに50言語のツリーバンクが公開されている。高精度な構文解析の実現にはツリーバンクが必要不可欠であることから、今後は大規模かつ高品質なツリーバンクを低コストで構築する技術などが必要となるだろう。

これまでの構文解析技術は、文の表層的な手がかり（品詞、機能語、接頭・接尾辞など）を利用して正しい構文木を選択している。一方、人間は表層的な手がかりに加えて、意味的な手がかり、あるいは意味理解をしながら整合性のある解釈を選択していると考えられる。このような、より人間らしい構文解析を目指す試みはたびたび行われているものの、今のところ全くうまくいっていない。

つまり、現在の構文解析技術は表層的特徴の統計的性質に過度に依存しており、学習データとは統計的性質が異なるテキスト（文のスタイルが大きく変わる、専門的文書で未知語が多いなど）では、解析精度が大幅に低下することが知られている。これまでの構文解析研究は、実用的には成功を収めつつあるものの、人間の言語理解の本質にはまだ迫ることができていないということもできる。

1.4.2.2 意味構造解析

構文解析の次のステップとして、文が表す意味構造を認識する技術を「意味構造解析」（semantic parsing）という。意味構造としては、述語とその意味的な項（意味的な主語や目的語）を表す「述語項構造」（predicate argument structure）や、述語論理などの形式論理を用いて意味を記述する「論



$$\exists x,y,e.(taro(x) \wedge mary(y) \wedge ask(e, x, y, \forall z.\exists f.(dog(z) \supset chase(f, y, z))))$$

■図21 「Taro asked Mary to chase all dogs.」に対する述語項構造（上）と論理表現（下）

※1
コーパスとは、自然言語処理の研究のために大量に集積し、構造を解析した情報も付加された言語資料のこと。

理表現」（logical form）（**図21**）、汎用的意味表現として提案された「AMR」（Abstract Meaning Representation）が挙げられる。

意味構造解析には二つの流派がある。一つは、構文木（句構造や依存構造）を入力として、意味構造の各要素を推定する機械学習分類器を学習する方法である。例えば、述語項構造は**図21**に示すように単語間のグラフ構造とみなすことができる。したがって、単語の各ペアについて、述語－項関係が成り立つかどうかを分類すればよい。これは、各単語について述語との関係（意味役割という）を認識する問題に帰着できるので、意味役割付与とも呼ばれる。単純な分類問題なので、SVM、パーセプトロン、CRF、RNNなど、様々な分類器や系列ラベリング手法を応用することができる。

もう一つの方法は、構文木に沿って意味構造を組み立てる方法である。**図21**に示した論理表現は、グラフ構造ではなく述語論理式であるため、上述のような分類問題に帰着するのは難しい。そこで、各単語に意味表現の断片を割り当て、構文木にそってこれを合成していくことで、文全体の意味表現を計算する。これはモンタギュー文法[2]から続く伝統的なアイディアであるが、文法理論に基づく構文解析が実用レベルに達したことと、形式論理に基づく意味表現の理論的進展から、最近では高精度な意味解析が実現されつつある。

構文解析と同様に、意味構造解析の研究においても言語リソースが重要な役割を果たしている。英語では「Proposition Bank」や「FrameNet」、日本語においては「NAISTテキストコーパス」といった述語項構造コーパスが開発されており、意味構造解析の研究をリードしている。ただし、意味構造のコーパスの開発はツリーバンクより更に困難であるため、このようなリソースが利用できる言語は限られている。

ここで述べた技術は、一つの文の意味構造を認識することを目的としている。しかし、実際の文章では、意味構造に必要な全ての情報が一文で完結することはなく、しばしば先行する文で言及された単語やフレーズが参照される。完全な意味表現を得るためには、文外の情報を適宜参照する必要があり、共参照解析・照応解析として研究が行われている。

1.4.2.3 グラウンドされた意味構造解析

これまで述べた意味構造解析は、自然言語文の意味を表す汎用的な表現を求めることを目的としていた。一方、出力すべき表現がアプリケーションから要請されるケースもある。例えば、大規模データベースである「Freebase」[3]に対して自然言語で問い合せをしたいとする。この場合、Freebaseからデータを取り出すためにはSPARQL言語[4]でクエリ[5]を書かなければならない。したがって、入力の自然言語文を、最終的にはSPARQLクエリに変換することが要求される。これは、自然言語の構造解析の立場からすると、SPARQLクエリを意味表現とみなせば、意味構造解析の一種と見ることができる。

このように、自然言語文をデータベースクエリなどのアプリケーション依存の形式言語に翻訳するタスクを、狭義の意味構造解析、あるいはグラウンドされた

■表1　Grounded semantic parsingの代表的なリソース

GeoQuery	米国の地理に関する質問応答
Jobs	求職情報に関する質問応答
Free917	Freebase に対する質問応答
WebQuestions	Freebase に対する質問応答

※2
米国の論理学者リチャード・モンターギュ（Richard Montague）氏が1973年の論文で記号理論における意味論と自然言語における意味論には本質的な違いはないとしたことに始まる、自然言語を記号理論の論理式で表現し構造解析を行う意味解析のアプローチ方法。

※3
Google（米国）が提供する知識情報データベース。

※4
SPARQL（SPARQL Protocol and RDF Query Language）言語は、ウェブ上にあるリソースの関係を記述するRDF（Resource Description Framework）に対して検索をする際のクエリ言語。

※5
検索内容を式として表したもの。検索式。

意味構造解析（grounded semantic parsing）という。**表1**に示すように、これまでにいくつかのデータセットが開発されている。

基本的なアプローチは、前述の意味構造解析と大きく変わらない。句構造を用いる手法、依存構造を用いる手法、CCG（Combinatory Categorial Grammar）を用いる手法など、様々な手法が提案されている。ただし、このタスクでは、学習データとして入力文と意味表現だけが与えられ、中間の構文構造は与えられないことが多い。したがって、完全な教師あり学習を用いることができず、構文構造を隠れ状態とした弱教師あり学習[6]が適用される。

つまり、正解の意味構造を導く構文構造は分からないため、構文構造を列挙して意味表現への写像を計算し、意味表現が正しいかどうかを判定基準として、構文構造と意味構造の曖昧性解消[7]モデルの学習を行う。更にこの考え方を拡張して、正解の意味構造を与えず、質問に対する答えだけから意味構造解析モデルを学習する手法も提案されている。この場合は意味構造も隠れ状態として扱われ、正しい答えを得られるかどうかという教師信号を利用して学習を行うことになる。構文木や意味構造を教師データとして与える場合より難しい問題設定であるが、高精度な意味構造解析モデルが学習されることが示されている。

1.4.3 自然言語の意味・知識理解技術

1.4.3.1 意味解析

1.4.2項で述べた構造解析技術は、文の構造的情報、すなわち単語間の依存関係を明らかにするものである。自然言語処理の分野では、1990年代後半から、更に単語の意味に深く踏み込んだ解析（意味解析）が行われるようになった。しかし、具体的に「意味」とは何か？　どのような「意味」が解析できればよいか？　計算機が扱える形で意味解析の問題を定義するには？　人間にもある程度解けるような問題か？　意味解析の研究では、基礎解析以上に、意味解析の問題をどのように“設計する”かが重要である。

このような問いに答える形で、多くの研究者により多種多様な意味解析問題の設計と評価用データセットの構築が行われた。また、MUC（Message Understanding Conference）、SemEval（Semantic Evaluation）、CoNLL（Conference on Natural Language Learning）Shared Taskといった評価型ワークショップ[8]が開催され、意味解析問題の標準化・データの共有が行われた。その全てを本書で紹介することはできないため、いくつかの重要な問題を紹介する。

「感情極性解析」（Sentiment Polarity Analysis）は、ある文、又は文章が与えられたとき、そこに表明されている意見がポジティブなものか、ネガティブなものかを解析する技術である。例えば、レストランのレビュー記事に書かれた「The atmosphere was good. I really liked the cesar salad.」（雰囲気は良かった。シーザーサラダは本当に気に入った）に対して、「ポジティブ」を出力する。意味解析タスクの評価型ワークショップであるSemEvalにおいても、過去6回共通タスクとして採用されており、実に多くの解析手法が提案されている。

「含意関係認識」（Recognizing Textual Entailment; RTE）は、ある2文、TとHが与えられたとき、TがHを含意するか（Tが真のとき、Hも必然的に真となるか）を解析する問題である。例えば、下記の

※6
正解の一部が与えられ、残りの正解の推定を含め学習を進める方法。

※7
一つの文に対して文法的には複数の構文構造が成立可能であり、その中から人が実際に解釈する構文構造を選択すること。

※8
「評価型ワークショップ」とは参加者に共通のデータを提供し、そのデータの評価（意味解析）を競うコンテスト型のワークショップ。

2文が与えられたとする[2]。

T: Cavern Club sessions paid the Beatles £15 evenings and £5 lunchtime.
H: The Beatles perform at Cavern Club at lunchtime.

一般的に、Cavern Club sessionsがthe Beatlesに対してlunchtimeに£5を支払った（paid）ということから、The BeatlesがCavern Clubで演奏した（perform）ということが推論できる。つまり、この例に対しては、「含意する」を出力する。含意関係認識の問題は、バルイラン大学（イスラエル）のイド・ダガン（Ido Dagan）氏らの研究グループの主導により、2006年に「Pascal RTE Challenge」として共通タスク化され[2]、過去7回の評価型ワークショップが行われた。最近では、2文の「含意」関係でなく「類似」関係の判定をタスクとする、STS（Semantic Textual Similarity）というタスクも提案され、2012年からSemEvalのタスクの一つとして採用されている。

意味解析問題の標準化、評価データの整備が行われると、意味解析手法の研究も大きく進展した。多くの意味解析問題は古典的な分類問題・構造予測問題に帰着されるため（例えば、感情極性解析と含意関係認識は二値分類問題である）、解析モデルそのものは、機械学習分野で古くから研究されているSVM、CRFといった基本的なものが用いられた。

前述の2例からも分かるように、意味解析の問題を解く上で重要なのは、人間が持つある種の常識的な知識を解析モデルに取り込むことである。先の例で言えば、「バンドに演奏をしてもらうためにはお金を支払う必要がある」「The Beatlesはバンドである」といった知識である。こうした、人手により書き尽くすことが困難で、小規模な評価用データから学習することも困難な知識を、いかにして解析モデルに組み込むか、といった観点から多くの研究が行われている。

1.4.3.2 文脈解析

1.4.2項で述べた自然言語の基礎解析技術は、一つの文を解析の対象とした技術である。しかし、我々が日常的に目にする本、新聞、ブログなどに存在する自然言語は、文のまとまり、すなわち文章である。一般に、文章全体の意味を考慮した自然言語の処理は「文脈解析」と呼ばれ、多くの研究が行われてきた。

代表的な例として、「照応解析」がある。照応解析は、文章内に存在する照応表現（代名詞など）について、その指し先を明らかにする解析である。例えば、「John shouted at Bob. He was angry.」という文章が与えられたとき、Heが指し示すものはJohnである、ということを同定する。解析手法については意味解析と同様、基本的な機械学習モデルに基づいて、性別の一致や同義性認識といった意味的な特徴量をいかに設計するか、といった点で多くの研究がなされ、今では解析器がソフトウェアとして公開されて広く使われるなど、限られた場面においては実用段階に達しつつある。

また、省略された代名詞に対する照応解析はゼロ照応解析と呼ばれ、代名詞が頻繁に省略される言語（中国語、日本語など）では、特に大きな問題となっている。ゼロ照応解析では、代名詞の情報（性別、物・人の区別など）を解析に用いることができないため、特に難易度の高い問題として広く認識されている。例えば日本語では、その最高性能は3割程度にとどまっている[3]。

談話構造解析は、文章内の文間の意味的構造を明らかにするタスクである。例えば、「John shouted at Bob. He was angry.」という文章では、2文目が1文目の「理由」となっていることを同定する。「構造」の仕様については様々な選択肢が考えられ、実際にこれまでに多くの議論があるが、実際にはRST（Rhetorical Structure Theory）などが多く用いられている。また、今日、談話構造解析の研究において最もよく用いられている評価データとして、ペンシルベニア大学が主導する「Penn Discourse Tree Bank」（PDTB）がある。PDTBは、ニュース記事の2文間に対して、約3万の意味的関係を付与したも

ので、談話構造が付与されたデータセットとしては大規模なものである。2006年のデータ公開を皮切りに、様々な談話構造解析手法の研究がPDTBを用いて盛んに行われている。

1.4.3.3 ディープラーニング・分散表現の登場

「Googleの猫」に代表される、2010年代に起こったディープラーニングによるブレークスルーは、自然言語処理の研究にも大きな影響を与えた。一つの大きなブレークスルーは、言語の分散表現（Distributed Representation、又はEmbedding）[9]学習の成功であろう。最も代表的な枠組みは、Googleのトマス・ミコロフ（Tomas Mikolov）氏によって提案された、通称「Word2Vec」[4]である。文献[4]は、分布仮説（Distributional Hypothesis）の分散表現版ともいえるアイディアにより、大規模コーパスから単語の表現学習が効果的に行えることを示した。例えば、学習されたベクトルの足し算・引き算により、ある種の意味的な演算が行えることを示した（例えば、V（King）−V（Male）+V（Female）≒V（Queen）など）。なお、自然言語処理における分散表現に関する研究については、文献[5]を参照されたい。

このほか、前述した基礎解析技術を含め、意味解析、文脈解析といった、あらゆる自然言語処理の研究においてもディープラーニング化の波が押し寄せた。例えば、畳み込みニューラルネットワーク（CNN）を利用した感情極性解析、RNNに基づく含意関係認識、RNNに基づく共参照解析、FFNNに基づく談話構造解析などが提案されている。しかし、その効果は画像や音声分野ほどのインパクトがなく、性能の向上幅は極めて限定的であり、依然として残された課題は多い。

画像や音声など、意味が詰まった、又は自己完結した“信号的入力”とは異なり、言語は人々が共有している知識を呼び起こすだけの、いわばトリガーのようなものでしかない。つまり、言語はまさに「記号」なのである。ディープラーニングの特徴は特徴量の自動学習であるが、こうした人の知識に依拠した「記号」の列から意味のある特徴量を取り出すことは、画像や音声とは異なる難しさがあると考えられる。それゆえに、人が持つ知識をいかに獲得・モデル化し、それらを解析の中でうまく使いこなせる計算機構をいかに作るかが、依然として本質的で重要、かつ困難な課題であり、未解決な問題として残されている。特に文脈解析といった高次の解析になるほど、この傾向は顕著であると考えられ重要となる。

1.4.3.4 議論マイニング

2010年代、議論マイニング（Argumentation Mining）と呼ばれる研究コミュニティに大きな動きがあった。議論マイニングとは、小論文や学術文献などの「議論」に関する文章を主な対象とし、情報抽出・談話構造解析を行う研究分野の一種である。2000年代においては、自然言語処理とは異なる文脈で研究が行われ、2006年に計算機上での議論モデルに関する国際会議Computational Models of Argument（COMMA）が初めて開催され（隔年開催）、2014年には自然言語処理のトップ会議であるACL（Association for Computational Linguistics）において、その第一回ワークショップWorkshop on Argumentation Miningが開催された（第四回目は2017年、自然言語処理のトップ会議EMNLP（Conference on Empirical Methods on Natural Language Processing）で開催予定）。

議論マイニングの研究は、意味解析・文脈解析の研究と同様、問題設計に関する議論や、評価用データの構築に関して数多くの議論がなされるところから始まり、今まさにこれを解くための計算機モデルの検討が始まったばかりである。代表的な問題設定として、ある議論のトピックに対する著者の立場（賛成や反対など）を同定するスタンス認識（Stance Detection）、オンライン掲示板やディベートなどに

※9
「言語の分散表現」とは単語を高次元のベクトルで表現する方法。

おける複数の発言者の発言の関係を、支持（support）、反論（attack又はrebuttal）に分類する議論関係分類（Argumentative Relation Detection）、議論の各文をコンポーネント（「背景」「関連研究」など）に分類する議論ゾーニング（Argumentative Zoning）といった問題が提案されている。解析モデルとしては、これまでの意味解析と同様、単語などの表層情報や意味的特徴量に基づく機械学習モデルが用いられているが、その性能は十分なものでなく、検討の余地は多い。

また、現状の議論マイニングで取り組まれている問題は、基本的には古典的な談話構造解析の問題の延長であるが、これは議論の自動評価（Automated Essay Scoring）などの実際に想定された応用技術を要する入力に対し、少々のギャップがある。例えば、議論の自動評価では、文間が「証拠」の関係にあるということ以上に、書き手がどのような事実や類推を組み合わせて証拠と考えるに至ったのか、といったより深い議論の解析結果が必要とされる。今後の議論マイニングの研究では、こうした深い解析に取り組む研究も登場してくるだろう。

1.4.3.5 常識推論

これまでに紹介した意味解析、文脈解析の研究は、自然言語の解析の上で解消すべき言語現象を駆動力として発展してきた。一方で、近年「言語の理解とは何か」を出発点として、学際的に行われる自然言語処理の研究が一つの大きな流れを作りつつある。本項では、その最新動向を紹介する。

チューリングテストは、「機械が知能を持つか」を判定するテストとして古くから用いられているテストである。審査対象の機械は、審査員から見えない別室に置かれ、審査員はその機械とひととおりの会話を交わす。審査員がその相手を機械だと見破れなければ、その機械は知能を持つと判定されるのである。こうした機械の知能テストを目的として、2000年代後半から今に至るまで、いくつかのテストが提案され、常識推論（Commonsense Reasoning）タスクとして、大きな注目を浴びつつある。

南カリフォルニア大学のアンドリュー・ゴードン（Andrew Gordon）氏の研究グループでは、「知能を持つ」ということを事象の因果関係の予測能力になぞらえ、「COPA」（Choice of Plausible Alternatives）という常識推論問題を提案した。COPAは、前提（Premise）と二つの文Alternative 1、Alternative 2が与えられたとき、Premiseの結果（又は原因）として相応しい文を選ぶ問題である。例えば、下記の問題を見てみよう。

Premise: The man broke his toe. What was the CAUSE of this?
Alternative 1: He got a hole in his sock.
Alternative 2: He dropped a hammer on his foot.

前提Premiseにおけるつま先（toe）を怪我した（broke）ことの原因としては、靴下（sock）に穴が空いたから（got a hole）ではなく、金づち（hammer）を足（foot）の上に落とした（dropped）から、ということがより相応しい。つまり、正解はAlternative 2である。このほか、著者のウェブサイト[10]において、データセット1,000問が一般公開されている。

また、2016年、ロチェスター大学（米国）の研究グループは、COPAを拡張した「ROC Stories」という問題を提案し、10万ストーリーからなるデータセットを一般公開している。ROC Storiesでは、4文からなるストーリーが与えられたとき、そのエンディングとして最も適切な文を二つの選択肢から選ぶことを要求される。2017年1月には、ROC Storiesを対象としたコンペティションが開かれ、2017

※10
Choice of Plausible Alternatives (COPA) Website
http://people.ict.usc.edu/~gordon/copa.html>

年4月に自然言語処理のトップ会議の一つであるEACL（European Chapter of the Association for Computational Linguistics）のワークショップとして、各種システムと関連研究の発表が行われた。

この種の問題を解くには、計算機も因果関係に関する常識的知識を持っていなければならない。例えば、上の例では、「足にhammerを落とすと、怪我をする」ということを知っていなければならない。現状行われている研究の主な解法は、因果関係を表すキーワード（“because”など）や照応関係などの手がかりを用いて、大規模な文章の集合から常識的な知識を獲得し、これらを基に2文間の因果関係を統計的に計算する手法である。また、獲得した知識をSequence to Sequence学習モデルに投入し、ストーリーの生成器を構築するアプローチもある。COPA、ROC Storiesともに、まだ7割程度の精度でしか解析ができておらず（2017年4月時点）、これからの発展が楽しみな分野である。

また、ニューヨーク大学のアーネスト・デイヴィス（Ernest Davis）氏とトロント大学（カナダ）のヘクター・レヴェック（Hector Levesque）氏の研究グループは、「統語的手がかり（統語的役割や、述語の選択選好性など）だけでは解けない照応解析の問題が解けること」を基準として知能テストを定式化した[11]。このテストは、AI研究者テリー・ウィノグラード（Terry Winograd）氏にちなんで、「WSC」（Winograd Schema Challenge）と名付けられた。

テストの例を下記に示す。

（1）The city councilmen refused the demonstrators a permit because they feared violence.
（2）The city councilmen refused the demonstrators a permit because they advocated violence.

ここでは、theyの指し先を、（1）ではthe city councilmen、（2）ではthe demonstratorsと正しく同定する必要がある。このためには、COPAやROC Storiesと同様、「ある人が何かをfearすると、refuseする」といった常識的な因果関係の知識に基づいた予測モデルを構築する必要がある。（1）と同時に（2）のような問題が含まれているため、「主語が先行詞として選択されやすい」といった統語的な手がかりだけでは正しく解析ができないようになっており、常識的な知識を使いこなして初めて解ける問題集となっている点がポイントだ。

2016年には、AIのトップ会議であるIJCAI（International Joint Conference on Artificial Intelligence）[12]のワークショップとして、WSC（World Sudoku Championship）の第一回コンペティションが開かれた。コンペティションにおける最高性能は、大規模ウェブデータから獲得した因果関係知識とニューラルネットワークを組み合わせたモデルであったが、その性能は5割程度であり、まだまだ発展途上の段階である。こうした常識的知識を使いこなす能力は、AI研究の発展には欠かせないものであり、今後の注目分野である。

1.4.3.6 知識獲得

これまで見てきたように、意味解析、文脈解析、常識推論では、基礎解析以上に常識的な知識の利活用が重要である。この重要性は古く1980年代から認識されており、知識表現（Knowledge Representation）、知識ベースの構築、推論の枠組みに関する様々な研究がなされてきた。研究の初期段階は、これらの知識を人手により整備する試み（例えば、プリンストン大学のWordNet、カリフォルニア大学バークレー校のFrameNetが主流であったが、2000年代にいわゆる情報爆発の流れを受け、ウェブ上の

※11
“The Winograd Schema Challenge,” NYU Computer Science Website <http://www.cs.nyu.edu/faculty/davise/papers/WinogradSchemas/WS.Html>

※12
米国AI学会が主催している。

大規模な文章集合から常識的知識を獲得する研究が急速に発展した。

獲得された知識は、固有名詞に関する知識（locateAt）、オントロジー的な知識（animal-catなどの上位下位関係）、事象間関係知識（hungry-eatなど）まで多岐に渡る。本項では、近年大きな動きがあった、固有名詞の関係知識の獲得と、事象間関係知識の獲得の研究の最新動向を紹介する。基本的には、1.4.3項で述べた分散表現に基づく知識表現・推論がトレンドである。

「知識グラフ」（Knowledge Graph）とは、主に固有名詞の関係予測問題の文脈で発展してきた知識表現の一種である。その名のとおり、概念（ここでは固有名詞）をグラフにおけるノード（頂点）に、関係をグラフにおけるノードをつなぐエッジ（辺）としたグラフであり、固有名詞間の関係ネットワークを表現する。

初期（2000年代）の知識グラフの獲得手法としては、ワシントン大学のオレン・エツィオーニ（Oren Etzioni）氏の研究グループによって開発されたText Runner、ReVerbに代表されるような、テキスト中に記述された関係知識に対する半教師ありの分類器に基づいて、関係抽出を行うことで知識グラフを構築していた。だが近年は、予測のロバスト性向上のために、分散ベクトル表現を用いた知識グラフの表現が主流である。

もう一つの最近の知識獲得研究のトレンドとして、「スクリプト的知識」（Script Knowledge）に代表される事象間関係知識の研究がある。スクリプト的知識とは、ロジャー・シャンク（Roger Schank）氏が提唱した概念であり、同時に起こりうる典型的な事象の集合を表す。例えば、「レストラン」のスクリプトには、「椅子に座る」「メニューを見る」「注文する」といった事象の集合が含まれる。典型的なスクリプト的知識の自動獲得の手法は、“and then”などの語彙統語パターンや、照応関係を用いて事象間の関係知識をウェブコーパス[13]から大規模に獲得し、これらに基づいて因果の強さを統計的に推定するものである。

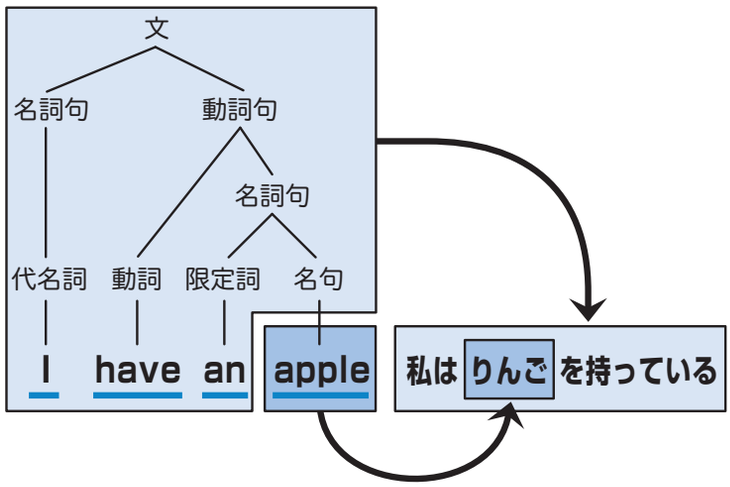
こうした自動獲得のアプローチにおける大きな課題の一つとして、コーパスから大量に獲得した因果関係の事例をどのように一般化し、知識とするか、という問題がある。例えば、「John was fined because he smoked in a non-smoking hotel room.」、「Mary smoked in a non-smoking room, so she was fined 10,000 yen.」といった因果関係の事例からは、“人がnon-smoking roomで喫煙すると、罰金を払わされる（fined）”という一般的な知識が得られる。しかし、これを更に一般化した“人が喫煙すると、罰金を払わされる”というのは（少なくとも相対的に）不適切に見える。

この問題を解消するために、近年はニューラルネットワークに基づくスクリプト知識のモデル化の研究が盛んに行われている。従来の研究では、一般化の粒度を「動詞のみ」「主語、動詞、目的語」などに固定していたのに対し、分散表現に基づく手法では、できるだけ事象に関する多くの情報を入力し、因果関係の推定が正しく行えるような事象の表現（すなわち、一般化の粒度ともいえる）を自動的に学習する、ということが期待される。ほかに、Word2Vec[4]を因果関係記述の集合に適用し、因果関係を予測するのに特化した分散表現（Causal Embedding）を学習する手法、獲得した因果関係を一般化した上で知識グラフ（ノードが事象、エッジが因果関係となる）を構築する手法がある。こうした研究によって、意味解析・文脈解析における長年の課題であった、常識的な知識の効果的な利用方法について、展望が見え始めている。

※13
ウェブ上から利用できる、言語の文章を大規模に集めたデータベース（コーパス）。

1.4.4 自然言語の生成技術

コンピュータが得た知識を自然言語によって表現し、テキストや音声の形で提示するという考えは、人間をユーザとするAIシステムの出力手段としては極めて自然なものの一つであろう。このような自然言語生成に関する技術は、自然言語処理の主要な目的の一つとして研究されており、様々な手法が提案されてきた。



■図22　構文情報を用いた伝統的な機械翻訳の例(Tree to string統計的機械翻訳)

1.4.4.1 具体的な構造に基づく自然言語生成技術

伝統的には、入力された知識情報と生成対象の自然言語との部分的な対応関係を列挙し、これらの断片的な情報に推論や探索アルゴリズムを適用することで、出力文として最も整合性の高い組合せを探索する手法が長年研究されてきた。これらの手法で特徴的なのは、入出力の具体的な記号同士の対応関係を明確にし、考慮する点である。

例えば、**図22**に示すのはTree to string統計的機械翻訳と呼ばれる伝統的な機械翻訳システムの動作例であるが、英語と日本語の文の断片「I have an...」と「私は...を持っている」、「apple」と「りんご」といった個別の対応関係をシステムが直接データベース上に保持しており、入力「I have an apple」が与えられた際には、無数に考えられる断片の組合せの中から回答として相応しい選択肢を絞り込むことで出力文「私はりんごを持っている」を生成する。対話システムや質問応答システム等においても、システムによる推論で導き出された情報を、あらかじめ用意したテンプレートやルールの集合に合致するよう配置することで出力を生成する手法が用いられてきた。

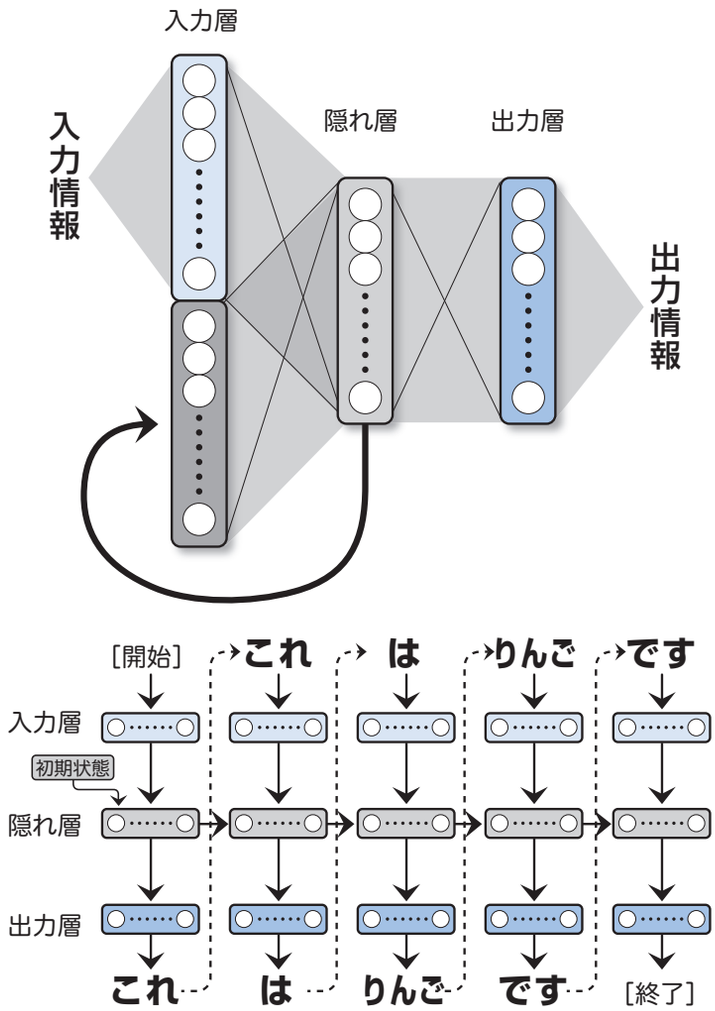
このような手法では入出力言語の統語構造や論理構造を知ることが大きな手がかりとなり、前項で紹介した構文解析技術、及び意味解析技術が重要な役割を担う。また文生成に用いられるデータ構造と実際の文の構造が密接に関連しているため、実際の処理を分解することで、どのような手順で出力文が生成されたのかを人間が明示的に知ることができる。このため、伝統的な文生成システムは技術者によって内部の挙動を直接観測・制御することが可能である。このような特徴は、実際のアプリケーションへ文生成システムを適用する際には利点となる。

一方、システム内部の挙動が複雑なデータ構造に支配されている点が、あるアプリケーションで使用された手法を他のアプリケーションにそのまま移植することを難しくする要因にもなっている。例えば機械翻訳システムの自然言語生成機能を他のアプリケーションに応用したい場合、アプリケーション側では機械翻訳に適したデータ構造を中間データとして出力する等の工夫が必要となり、この作業にしばしば多大な労力を必要とすることとなる。

1.4.4.2 リカレントニューラルネットワーク言語モデル型デコーダによる自然言語生成技術

人間の日常的な会話で発生する自然言語生成では、常に厳密な文の構造を意識しながら単語の選択を行っているとは考えにくい。話者が話そうとする情報に基づいて、前後の単語同士が流暢に接続するよう順に単語を選択しているととらえるのが妥当であろう。

このような挙動を実際の文生成システムに当てはめて考えると、入力データと初期状態を受け取り、出力文の単語を逐次的に出力するオートマトン[14]の一種としてとらえることができる。このオートマト



■図23　RNNの基本構造（上）
RNNを使用した自然言語生成の仕組み（RNNLM型デコーダ）（下）

ンに必要となるのは、現在のシステムの状態に基づき、出力単語列の言語としての流暢さを考慮した上で、次回の出力として最も妥当な単語を決定する仕組みである。

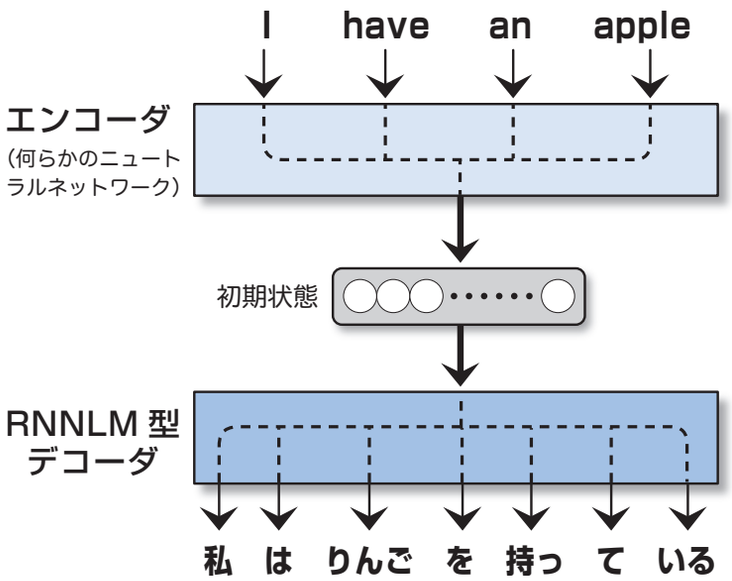
このような仕組みは言語モデルと呼ばれ、出力文の流暢性を担保する重要な素性として、様々な自然言語処理システムの構成要素として導入されてきた。言語モデルの本質は「過去の入力系列から次回の出力単語の分布を求める確率モデル」であり、単語の分布の定式化により、大きくn-gram言語モデルとリカレントニューラルネットワーク言語モデル（Recurrent Neural Network Language Model; RNNLM）の2種類に大別される。このうち、後述の理由により、近年の研究ではRNNLMを自然言語生成システムとして採用するものが増えつつある。

RNNLMは、**図23**のようにRNNを内部状態として採用した言語モデルである。同様のネットワーク構造は1980年代に既に提案されているが、これを言語モデルに応用したミコロフ氏らの研究は、ディープラーニングの手法を自然言語処理に流入させる契機となった点で、自然言語処理研究の転換点として重要な位置にある。

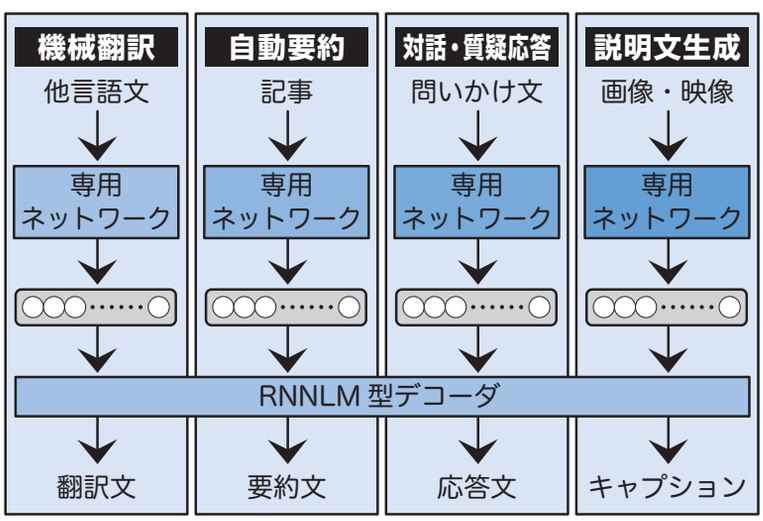
RNNLMの出力単語を入力にフィードバックさせることで、入力としてRNNの初期状態ベクトルの

※14
外部からの入力に対して、内部の状態に応じた処理を行い，結果を出力する処理のこと。同じ入力に対して、内部の状態によって異なる出力を行う場合があることが特徴である。

みを受け取り、文末まで自動的に単語を生成し続ける自然言語生成システム（デコーダ）を構築することができる。このような仕組みを本節ではRNNLM型デコーダと呼ぶこととする。



■図24　エンコーダ・デコーダモデル（機械翻訳の場合）



■図25　様々なタスクで共通して使用可能なRNNLM型デコーダ

1.4.4.3 エンコーダ・デコーダモデル

RNNLM型デコーダの初期状態ベクトルは、本質的には単なる実数ベクトルであり、自然言語の生成に必要な情報が格納されていること以外には基本的に値の制約が問われない。

つまり、入力情報を単一の実数ベクトルにエンコードする任意の手法（エンコーダ）を接続することが可能である。この特徴により、従来の具体的なデータ構造を入力データとして使用する自然言語生成システムに比べて、RNNLM型デコーダは様々な入力とモデル的に連携しやすいという特徴がある。

ここで、何らかのニューラルネットワークにより構成されたエンコーダをデコーダに接続することで、エンコーダとデコーダ双方を単一のニューラルネットワークと見なして同時に学習することが可能となる。このように、中間データを介してエンコーダとデコーダが対向する形で構成されたニューラルネットワークは、「エンコーダ・デコーダモデル」と呼ばれる。

最も初期のエンコーダ・デコーダモデルとして、エンコーダにもデコーダとほぼ同様のRNNを備えたエンコーダ・デコーダ機械翻訳モデルが挙げられる。これに注意機構を加えたモデルは、現在の機械翻訳システムの主流の構成法であり、2014年頃から著名な機械翻訳コンペティションにおいて次々と最高精度を達成し続けているのは記憶に新しい。

Googleが2016年に発表した機械翻訳システム「GNMT」（Google Neural Machine Translation）も、本質的にはエンコーダ・デコーダ機械翻訳モデルにいくつかの改良を加えたものである。**図24**では、エンコーダ・デコーダ機械翻訳モデルで先程の例「I have an apple」を「私はりんごを持っている」に翻訳する例を示している。まずエンコーダが英単語「I」「have」「an」「apple」を順に読み込み、中間データである初期状態ベクトルを生成する。デコーダは初期状態ベクトルを受け取り、自身の状態遷移にしたがって単語列「私」「は」「りんご」「を」「持っ」「て」「いる」を生成する。このモデルの場合、初期状態ベクトルが入力文に関する情報を全て蓄えていることは明らかである。

エンコーダ・デコーダ機械翻訳モデルは入出力がともに可変長の単語列であることから「Sequence to Sequenceモデル」とも呼ばれる。また、単一のニューラルネットワークで入力から出力までの全てを扱うモデルを「End to Endモデル」と称することがあり、エンコーダ・デコーダモデルはEnd to Endモデルのサブセットであるということもできる。

前述したように、RNNLM型デコーダを使用し、エンコーダ部分を入力データに適したネットワークに変更することで、様々なデータ形式から自然言語を生成するシステムを構築可能である（**図25**）。例えば、前述のように自然言語を受け付けるエンコーダを使用することで機械翻訳や自動要約システムの基本的な構成を作ることができ、CNN等の画像を認識可能なネットワークをエンコーダとすることで、キャプションや映画の字幕・あらすじの生成等を実現することができる。

実際の人間の脳活動と関連した研究としては、機能的MRI（functional Magnetic Resonance Imaging; fMRI）画像を認識するエンコーダを接続することで、脳活動の情報から自然言語を生成する試みがある。このように、従来モデル的に連携が難しかったデータ構造同士が、エンコーダ・デコーダモデルの枠組みの下で有機的に結合することが可能となった点は特筆に値する。

1.4.4.4 リカレントニューラルネットワーク言語モデル型デコーダによる自然言語生成の課題点

このように、様々な入力に対して有効性が期待されるRNNLM型デコーダ（及びエンコーダ・デコーダモデル）であるが、従来型の自然言語生成システムには存在しなかった特有の問題が複数存在することも報告されている。

特に、単純な構成のRNNLM型デコーダは初期状態を与えれば全自動で動作するため、外部からデコーダの挙動を制御する手段が存在しない点がしばしば問題となる。これにより引き起こされる主な問題として、「過生成」「生成不足」の2点がある。

過生成とはデコーダの状態遷移が何らかの不動点に陥ってしまい、同じ情報を繰り返し（しばしば永遠に）出力し続けることをいう。生成不足とは逆に、本来必要な状態遷移の一部をスキップしてしまったがために、入力として与えられた情報の一部のみが自然言語として出力されてしまうことをいう。**表2**は機械翻訳における過生成と生成不足の典型例である。このような問題がどのような条件で発生するのかは今のところ詳しい解析がなされていない。モデルの構造を工夫することによって間接的に回避するための手法も研究されているが、数理的に解決されたとは言いがたく、今後の更なる解析が待たれる。

また、RNNLM型デコーダは本質的には単なる言語モデルの一種であるため、生成される自然言語の流暢さのみが過剰に追求され、本来の入力データの情報がデコーダにより改変されてしまう場合がある。この問題はGoogleによってGNMTがアプリケーションとして公開された当初から、翻訳結果の統語構造が入力文とは異なっていたり、入力文に全く存在しない名詞が突如として出現したり等、具体的な現象として指摘されている。

特に、言語モデルの効果によって出力文単体としては非常に流暢な文が得られるため、入力された正しい情報を知る手段を持たない者（機械翻訳であれば、入力言語を理解し得ない者）がデコーダの出力に「騙されて」しまう可能性があり、RNNLM型デコーダを使用する上で無視できない問題である。

なお、RNNLM型デコーダ単体としては文の意味合いを明示的に保存するような仕組みはなく、単に入力された情報と統計

■表2　RNNLM型デコーダにおける典型的な誤動作の例

過生成	入力：I have an apple and an orange 出力：私はりんごと**りんごとりんご**とオレンジを持っている
生成不足	入力：I have an apple and an orange 出力：私はりんごを持っている（**オレンジを省略**）
原構造無視	入力：I have an apple ー pen 出力：私は**りんごとペン**を持っている

的に合致する単語列を出力しているだけと考えるのが妥当であろう。このような問題が現状では見逃されている背景として、出力された自然言語文の評価法に問題があると考えられる。自然言語処理研究においてシステムの出力をどのように評価するかは重要な課題だが、現在の主な研究ではテストデータ全体に対する平均的な文の一致を計算する自動評価尺度（代表的な尺度として、機械翻訳ではBLEU（Bilingual Evaluation Understudy）、自動要約ではROUGE（Recall-Oriented Understudy for Gisting Evaluation）など）を使用する傾向が強い。

これらの尺度は文の特定の部分で大きな破綻をきたしているかまでは着目しないため、人間が文を理解するに当たって許容し難い誤りを過小評価してしまう傾向にある。これらの自動評価尺度は開発されて既に10年以上経過しているものが多く、この点でも現在主流の手法を正しく評価できるとは考えづらい。本来、問題解決のための手法とその評価法はともに進歩してゆく必要があるが、手法のみが単独で先行してしまっている状況にある。

対話システムや自動要約にRNNLM型デコーダを用いる場合は、デコーダが出力する情報の可制御性がより重要な問題となる。対話システムの場合、ユーザにシステムが提示する情報は、システムが生成する文に過不足なく表現されている必要がある。これを実現するために、対話制御のための内部状態をRNNの要素として導入することで一定の可制御性を担保する仕組みが提案されている。自動要約では生成される文の単語数が特に重要となるが、これを制御する機構を導入し、出力文の長さをユーザがある程度可能とした手法が提案されている。

一般的にニューラルネットワークの学習には大規模なコーパスが必要となるが、これはRNNLM型デコーダにおいても同様である。言語モデルや機械翻訳といったタスクでは、数百万文～数億文程度のコーパスが比較的容易に入手可能であるため、（種々の問題は残っているが）実用的なレベルの性能を達成可能なモデルを学習することができる。自然言語処理のなかでも、特にこれらの分野で最初にディープラーニングの応用が始まったのは、このような言語資源による背景も大きく影響していると考えられる。

一方で、話者が少数の言語や、対話システム等の大規模なデータ収集が難しい分野では、言語資源の少なさに起因するニューラルネットワークの適用の困難さが直接支障となる場合がある。このような問題への対処として、異なる言語・ドメイン上の大規模なコーパスで獲得されたモデルの流用や合成によってモデルの頑健性を向上させる試みがある。GNMTにおいて複数の言語を同時に学習させることで、未知の言語対においても翻訳可能なことを示した例はその最たるものであろう。これらは「教師データとして明示的に与えられない表現を学習する」という意味でゼロショット学習と呼ばれる傾向にある。

参考文献
[1] 黒橋禎夫·長尾眞「京都大学テキストコーパス・プロジェクト」『言語処理学会第3回年次大会予稿集』
[2] In Machine learning challenges. Evaluating predictive uncertainty, visual object classification, and recognising tectual entailment. I. Dagan, O. Glickman, and B. Magnini, “The PASCAL recognising textual entailment challenge,” Springer, pp.177-190.
[3] 大内啓樹ほか「文書全体を考慮したニューラル文間ゼロ照応解析モデル」『言語処理学会第23回年次大会発表論文集』pp.815-818.
[4] T. Mikolov et al., “Distributed representations of words and phrases and their compositionality,” Neural Information Processing Systems, pp.3111-3119.
[5] 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.