
付録 第1部

AI技術

第1章

AI技術 ～知的活動を実現する基礎技術～

AI技術 ～知的活動を実現する基礎技術～

本章は、「AI白書2020」の「2.2 知的活動を実現する技術」を再整理したものである。

1 認識

コンピューターの「認識(recognition)」とは、データの中に対象に関するパターンを見出す「パターン認識」処理といってもよい。データからある規則(パターン)を発見したり、規則(パターン)を使って、データがどのパターン=クラス*¹に属するかを分類し、判定したりする「クラス発見とクラス分けの問題」であると考えればわかりやすい*²。つまりデータの中に潜む特徴量を基に判定する技術である認識は、コンピューターの初期からの課題の一つである。人間の作業をコンピューターに代行させようとする、人間のように実世界を認識できる必要がある。このときの基本になるのが、人間の視覚や聴覚などに相当する機能だ。しかし、視覚や聴覚に対応する生データは、コンピューターからすれば背景が写り込むなど情報が過多であり、その中から対象に関連した特徴やパターンを抽出する手法が考えられてきた。

認識技術は認識対象に基づいて、「物体認識」「文字認識」「顔認識」「行動認識」「(自然)言語認識」などに分類される。また、認識するデータ構造に着目すれば「画像(静止画)認識」「映像(動画)認識」「音声認識」などに分類されることもある。認識の流れを図表(付)11-1に示す。

図表(付)11-1 認識の流れ



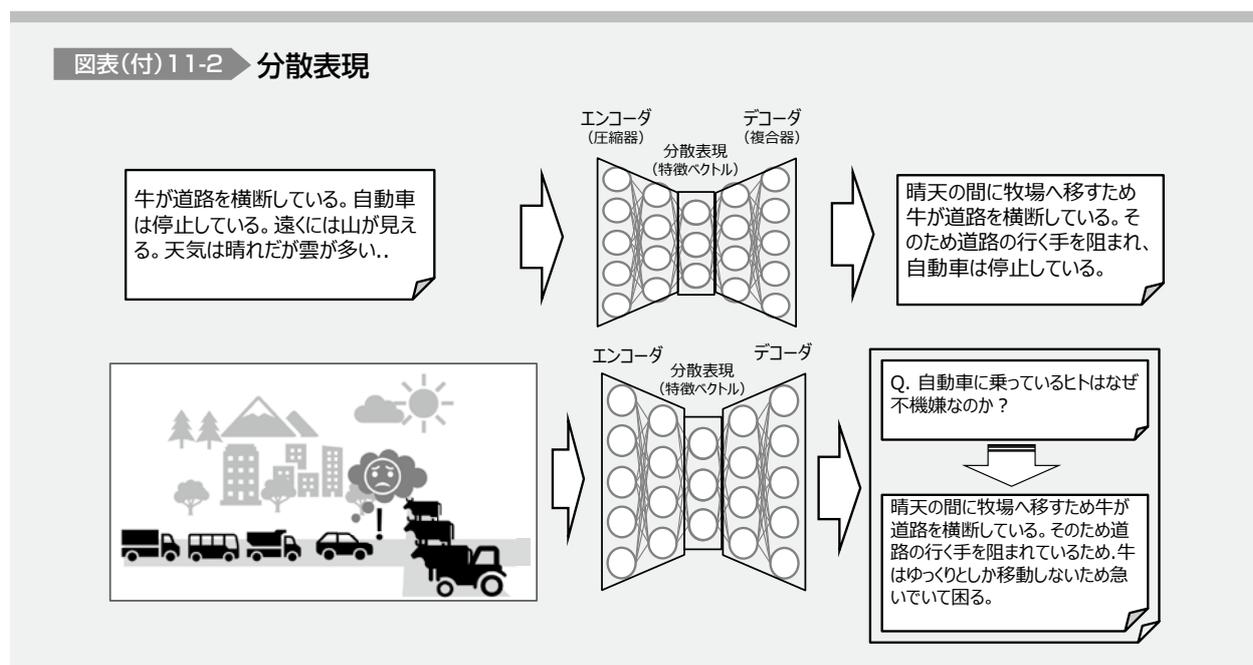
2 理解

「理解」とは「物事の道理や筋道が正しくわかること。意味・内容をのみこむこと。」「他人の気持ちや立場を察すること。」(デジタル大辞泉)である。しかし、人間同士であっても、自分が何かを理解できたか、相手が自分の気持ちを理解しているかを確認することは容易ではない。ましてや、コンピューターが何かを理解したか否かを外部から確認することは極めて困難である。そこでここでは、コンピューターが画像や言語を認識し、適切な応答をした場合に、コンピューターが「理解」したと表現する。とくに現在では特徴ベクトルによる分散表現をどのように扱えるかという点が重要な技術になっている。

* 1 クラスとは、たとえばネコやイヌや花などのカテゴリを指す。

* 2 M. Bishop (“Pattern Recognition and Machine Learning”, Springer, 2006)によると、「パターン認識の分野はデータから自動的に規則(regularities)をコンピューターのアルゴリズムを利用して発見することと、発見された規則を用いてデータを複数のカテゴリに分類するなどのアクションを起こすこと、に関連している」と定義される。

文章における単語の意味は、その単語の周辺の単語や文脈に依存し、コンピューターがこれらを考慮した応答ができれば意味を理解しているとみなせる。このような意味の理解に資する技術に分散表現がある。分散表現とは、簡単にいうと、与えられた文章において、単語が出現する周辺の単語の情報を使って、その文章における単語の特徴を多次元のベクトル(特徴ベクトル)で表したものである*3。分散表現の求め方の代表的なものに、大量の例文データを使い、ニューラルネットワークに分散表現を学習させる手法がある*4 (図表(付)11-2)。分散表現が意味を表すという根拠となるのは、意味の似ている単語は、似ている文脈(文章中の単語の並び)に出現するという「分布仮説」である。意味の似ている単語の分散表現は、ベクトルとして距離が近い分散表現に対応する。



ある画像に対して何が写っているのかをコンピューターに言語表現させたとき、ほとんどの場合に人間と同じと判定できるのであれば、コンピューターは画像の「意味」を理解しているとみなせる。そこで、画像を介して異なる言語で表現させることで、翻訳が可能になる。たとえば、日本語の「ネコ」という単語を使い画像を検索させ、表示された画像に対して英語で学習した画像認識を行わせることで“Cat”という英単語を得ることができる。さらに、さまざまな言語の単語に対して、同様の手順を繰り返すことで、画像を介して、複数の言語の単語を結びつけることが可能になり、画像の意味を理解したかのように見える*5。またVQA(Visual Question Answering)*6技術は、提示した画像の内容に関する質問に対して適切な回答を返す(例：バスの色は何色?→バスの色は赤色です)機能を深層ニューラルネットワークで実現し、コンピューターが画像の内容を理解しているかのように見せられることを示した。

* 3 ニューラルネットワークを用いて文章・単語を学習することで中間層に生成される一種のベクトル。
 * 4 情報を圧縮するエンコーダと復元するデコーダを組み合わせる場合に圧縮した状態を分散表現で使用する。
 * 5 Hewitt, John, et al. “Learning translations via images with a massively multilingual image dataset.” Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018.
 * 6 H. Gao et al., “Are You Talking to a Machine? Dataset and Methods for Multilingual Image Question Answering”, In Proc. of NIPS, 2015.

3 学習

コンピューターにおける「学習(Learning)」は、とくに「機械学習(Machine Learning; ML)」と呼ばれる。機械学習とは、データの背後に潜む規則性や特異性を発見することにより人間と同程度あるいはそれ以上の学習能力をコンピューターで実現しようとする技術^{*7}である。機械学習の処理は、大きく「学習」ステップと「推論」ステップに分かれる。データを利用してコンピューターに何らかの学習・推論を行わせる仕組みが「機械学習」となる。

「学習」とは、事前に与えられたデータからモデルをつくることをいい、「学習」に使われるデータをとくに「学習データ」と呼ぶ。学習が終わったモデル(「学習済みモデル」とも呼ばれる)には、学習データの統計的な傾向や規則性が反映されている。モデルの出力は、判定結果や、分類対象のクラスや、予測結果を示す数値になる。また出力の確かさ(確率)を出力する場合もある。

これに対して「推論」とは、実用環境で得られるデータなどを使い、学習済みモデルを使って判定、分類、予測などを行うことをいう。正確な推論を行うためには、学習時にデータから過不足のない特徴を抽出し、モデルを得る必要がある。

有用な学習モデルを作成するために機械学習では、代表的な三つの学習の枠組みがある。それぞれの学習の方法から「教師あり学習」「教師なし学習」および「強化学習」と呼ばれる(図表(付)11-3)。

「教師あり学習」とは、学習データに正解ラベルが付与されたデータを用いることで、正解ラベルに基づくデータを学習する。データと正解ラベルが質問と回答の関係、つまり教師の役目となるためこの名前がついている。たとえばネコとイヌを認識させるためにネコとイヌのラベルを付与した学習画像データを用いて教師あり学習を行う。学習後は、未知の画像がネコとイヌのどちらであるか、どちらにより近いかを推論できるようになる。これに対して「教師なし学習」は、正解ラベルを付与していない学習データからデータに内在する特徴量(構造、規則性等)を見つけ出すものである。たとえば、ラベルを付与していない大量のデータに対して、その中から何らかの特徴を見出してグループ分け(クラスタリング)を行うことができる。「強化学習」は、行動を選択可能なシステムに対して選択結果に沿った報酬^{*8}を与えることで、どの行動を選択すべきなのかという「方策」を学習させるものである。技術的な特徴を含めた詳細に関しては「AI白書2020」の「2.2.3 学習」を参照、確認されたい。

図表(付)11-3 機械学習手法の3分類

学習の種類	概要
教師あり学習	学習用データは正解値等のラベルが付与されている。このデータを利用して学習モデルを作成する。推論では、正解ラベルのない新たなデータに対して分類や予測を行う。応用には、自然言語処理、スパムメールフィルタリング、手書き文字認識、故障診断等がある。
教師なし学習	学習データには、ラベル等の情報は付与されていない。データそのものに隠されている構造や規則性を学習する。クラスタリングやデータ次元圧縮技術などがこれに含まれる。
強化学習	強化学習では、「エージェント」 ^{*9} が学習を行う中で、行動により変化する状態を観察し、あらかじめ決めた報酬(行動の良し悪し評価)を最大化するように試行錯誤しながら行動を選択し学習する。必ずしも、あらかじめ報酬を最大化する条件は決めない。システム自身が試行を繰り返すことで学習が進む。ロボットの動作学習に使われている。

* 7 JST(研究開発の俯瞰報告書)システム・情報科学技術分野(2019年):2.1.1 機械学習
<<https://www.jst.go.jp/crds/report/report02/CRDS-FY2018-FR-02.html>>

* 8 報酬とは、行動の評価、簡単には良し悪しに関する情報を指す。

* 9 環境とのインタラクションをしながらタスクを実行する機能。

4 判断

ここでは、「判断」とは、複雑な制約がある中で、何らかの評価関数^{*10}を用い意思決定をすることとする。意思決定の対象は、単独の行動や、組み合わせ(マッチング)や順番の決定(プランニング)である。

「意思決定」は、活動中は常に行われている。この仕事を誰に頼むか、昼食はどこで食べるか、気に入った商品を注文するかなど、さまざまな局面で意思決定を行う。業務であれば意思決定の誤りが、重大な結果に繋がることもある。そのため業務上の意思決定は、できるだけ多くの判断材料を集め、分析し、適切な意思決定を行わなければならない。インターネットを用いることで短時間に大量の判断材料を集めることが可能になったが、大量の情報が流れ込むことで、むしろ判断に迷うこともある。こうした状況で、大量の情報を基にAIが判断材料を提示し、「意思決定」の参考とすることは有用となる。また自動運転車やロボットのように、AI自身に「判断」を行わせるケースもある。

AIが判断を行う手法としては、第二次AIブーム^{*11}のときに知識処理として注目されたエキスパートシステムがある。専門的かつ高度な問題をエキスパートと同等のレベルで解決するための知的問題解決システムとなる。たとえば、専門家が意思決定する際には、法律や企業の内規、経験や嗜好からつくられた「こういう場合にはこうすべき」というルールを勧案するであろう。このようなルールをif-then形式で記述し、知識としてすべてをシステムに記憶させる。エキスパートシステムを使えば、たとえば、質問に対する答えとして知識に基づいた判断を得ることができる。

他の手法の例として、決定木分析が挙げられる。データを性質(例：男女)や実績値(例：販売数5個以上/以下)に基づき木構造で分析しておくことで、現状と照らし合わせて判断を行わせることができる。また、将棋や碁のように、現在の状況で打てる手の中でより効果が高い一手を探索していくこともできる。

これに対して、第三次AIブームで注目されているディープラーニング^{*12}は、膨大なデータを読み込み、適切な解を得られるようにモデルを訓練することで、従来の手法と比較して飛躍的に高い精度を実現している。将棋や碁においても、局面の画像を読み込むことで最適手を出力することができる。ただし、ディープラーニングによる判断は仕組み上、根拠を説明することが難しいため、実際に運用するには工夫が必要となる。たとえば、自動運転車については事故原因究明などのために「作動状態記録装置」の設置を義務化している^{*13}。またAIを用いた診断・治療支援を行うプログラムを利用して診療する場合については最終的な判断の責任を医師が負う方向での検討がされている^{*14}。AIによる判断のイメージを図表(付)11-4に示す。

* 10 評価対象である事象の良し悪しを数値化する関数。

* 11 AIブームはこれまでに3度あったとされており、たとえば総務省には簡単にまとめられた図がある。その中では第一次AIブームでは探索と推論、第二次AIブームでは知識表現、第三次AIブームでは機械学習が代表的なトピックと説明されている。

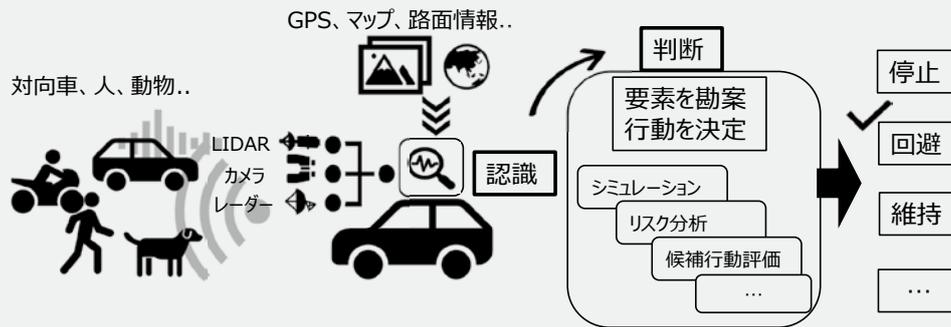
<<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/html/nc142120.html>>

* 12 多層の人工ニューロンによるネットワークを使用した学習技術。詳しくは「AI白書2020」の「2.3 ディープラーニング」を参照されたい。

* 13 <https://www.npa.go.jp/bureau/traffic/selfdriving/>

* 14 <https://www.mhlw.go.jp/content/10601000/000468141.pdf>

図表(付)11-4 AIによる判断



5 予測

予測とは、過去のデータから数学モデルなどを用いて未来のデータを求める処理をいう。AIでは、回帰分析や、そのほかの機械学習アルゴリズムを用いたモデル化による予測のほか、顧客を分類して購買特性から近々売れる商材を予測したり、データの変動パターンをクラスタリングして現状と照らし合わせることで今後の変動を予測したりするものも含まれる。予測が可能になれば、効率や成果の高い作業に資源を集中させることが可能になる。たとえば、従来の材料開発や薬剤開発では、俗に言う「しらみつぶし」な探索や試行が行われていた。しかし、最近では、AIなどの情報技術と材料科学を組み合わせたマテリアルズインフォマティクス(Materials Informatics)や生物学や医学と組み合わせたバイオインフォマティクス(Bioinformatics)が実用段階に入り、シミュレーションなどによる予測が可能になってきた^{*15}。従来は非常に難しいと考えられていた人体のほぼすべてのタンパク質の構造を予測することが可能になった^{*16}。

従来から行われてきた統計処理も機械学習も予測するための技術である。統計処理においては、そのために各種の統計値が使われていた。パーソナルコンピューターが普及し始めてから40年弱、高度な計算が可能な32ビット機が普及してから30年である。しかし、統計の源流は古代から始まっており、数理統計学にしても数百年以上の歴史を持つ。その大半の時間、人はデータをコンピューターなしで処理し、性質を「把握」していた事実がある。

これに対して機械学習は、最初からコンピューターでの利用を想定して、というよりも、むしろコンピューターを前提につくられた技術である。大量のデータに対して疲れることなく計算を続けることが前提となっている。現在主流の機械学習技術の一つであるディープラーニングでは、膨大なパラメーターを持つ多層モデルを用いることでデータに潜在する特徴の抽出を行う。学習を成功させるためには、数々の統計的・数学的・工学的な工夫を用いて、多層モデルのパラメーターが有効に働くように設計する必要はあるが、性能は非常に高い。反面、あまりにも多い学習パラメーターを解析しても得られた予測がどうしてそうなるのかといった説明が困難になるという問題が発生している。

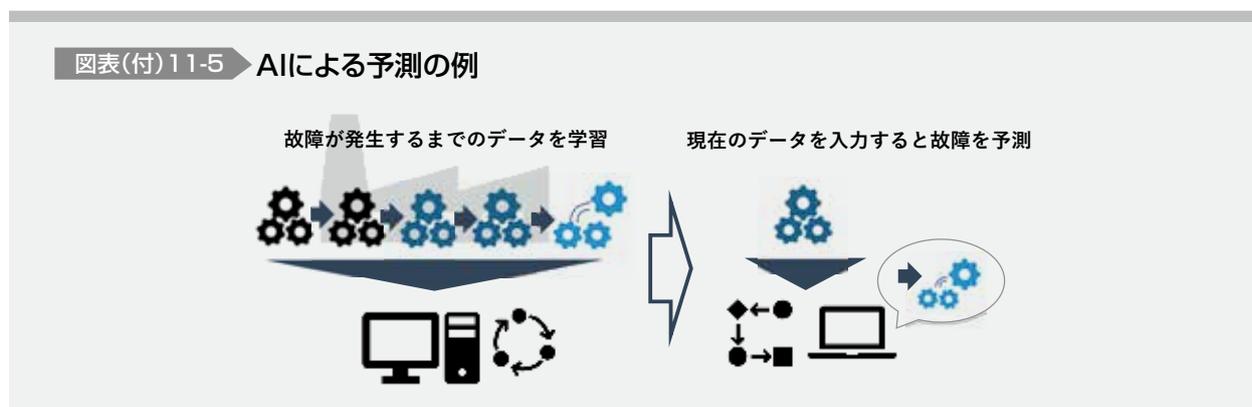
現在はディープラーニングとシミュレーションを組み合わせることで、大量のデータから得られた予測を生かしながら、予測の道筋を同時に解析し、説明可能な予測に結び付ける技術も進展している。

故障が発生するまでのデータを学習することにより、現在のデータから故障の予測を行い、その理由

* 15 例: <<https://xtech.nikkei.com/atcl/nxt/column/18/00001/05852/>>

* 16 <https://www.nature.com/articles/d41586-021-02025-4>

を説明する。数理統計学で使用する手法とディープラーニングの組み合わせにより、データに潜在する注目特徴量に焦点をあて、効率的な推定と説明を行う方法も試みられている。AIによる予測の例を図表(付)11-5に示す。技術の詳細は「AI白書2020」の「2.2.5 予測」を参照されたい。



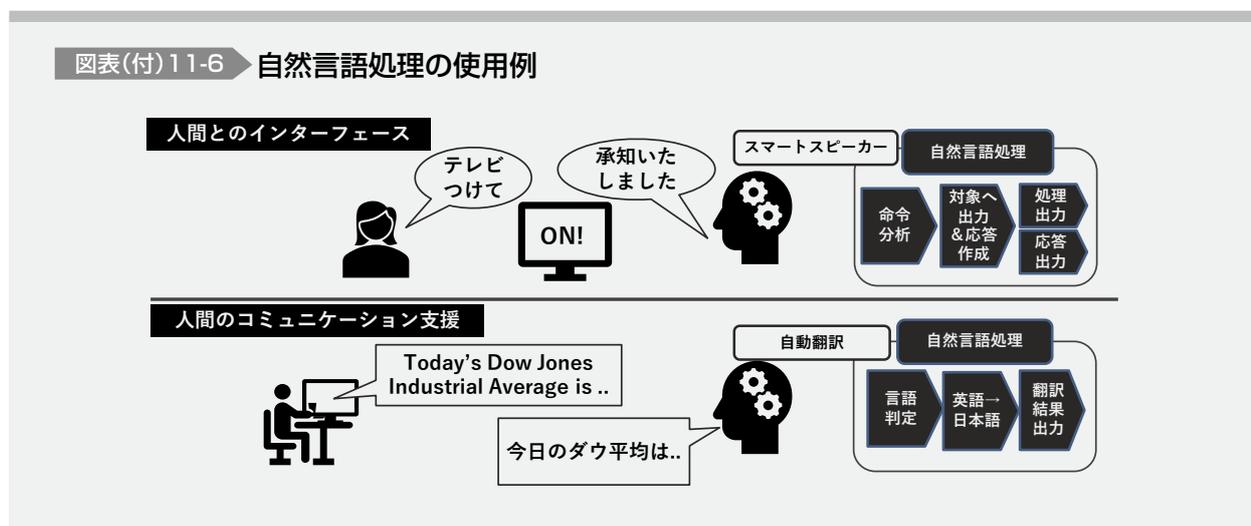
6 言語

コンピューターで自然言語を扱うことは、コンピューターの創成期から行われていた。また、コンピューター登場以前に言語学などで行われていたテキストの統計処理、テキストや発話を大規模に集積したデータベースであるコーパスの作成なども本来はコンピューター向きの作業であった。その中で機械翻訳は「自然言語処理最大のアプリケーション」*17といわれるように、コンピューターでの自然言語処理研究の中心にあった。しかし自然言語処理は、たとえば特定言語の単語数だけでも膨大なものがあり、コンピューター創成期のハードウェアには過大な処理でもあった。また自然言語は、言語学としてコンピューター登場以前から研究されてきたが、文法定義や品詞、意味分類は必ずしも一つに定まらず、コンピューター処理を想定したものではなかった。

1960年代から盛んになったワードプロセッサーなどの発展による文章の電子化に続き、「機械翻訳」は、多大な投資と、自然言語処理技術や計算機パワーの向上などにより、技術的な発展を遂げたが、最初に開発された、主に人手で整備するルールに基づいた機械翻訳(RBMT)は、その性能がすぐに頭打ちになった。1980年代には長尾らにより提案された用例に基づく機械翻訳(Example Based Machine Translation; EBMT)が提案された。翻訳の事例を大量に集めたコーパスを用いた翻訳手法である。高品質の大型両言語対訳コーパスが必要となり、実現は容易ではなかったものの数少ないこの分野での日本提案の技術である。IBMの研究グループが統計的機械翻訳(SMT)の研究を開始したのもほぼこの時期となる。これは単語の翻訳確率や並べ替えの確率などの翻訳に必要な知識情報提供の手段を対訳コーパスから統計的な情報として学習するものであり、現在でもスタンダードな機械翻訳手法として広く使われている。SMT手法の研究が頭打ちになったころ、全く新しい手法として2014年に登場したのが、ニューラルネットワークを用いた機械翻訳、いわゆるニューラル機械翻訳(NMT)である。深い(深層)ニューラルネットワークは、ここ数年で画像、音声、ロボティクス、自動運転などさまざまな分野において大きな成果を上げている。機械翻訳を含む自然言語処理も同様にその恩恵を受け、それまでの

* 17 西川 仁. “自然言語処理概論: 組合せ最適化の観点から(特集 自然言語処理と数理モデル).” オペレーションズ・リサーチ Communications of the Operations Research Society of Japan: 経営の科学 62.11 (2017) : 697-702.

成果を大きく上回る結果を残している。自然言語処理の使用例を図表(付)11-6に示す。言語の構造解析の詳細、機械翻訳の具体例については、「AI白書2019」の「2.3 自然言語処理」や「AI白書2020」の「2.2.6 言語」を参照されたい。



7 知識

「知識」はAIの歴史の中で中心的な役割を果たしてきた。1980年代には、「知は力なり」というかけ声の下、エキスパートシステムに代表されるように、if-then形式で表せるルール型の知識を収集して、人間の専門家の代わりとなるAIを実現する努力が進められた。しかし、集まった知識だけでは問題を解決できない場合が多く、下火になった。一方、1989年に始まったWWW (World Wide Web)はインターネット利用の起爆剤となり、急激に普及したが、他方、有象無象のビッグデータがインターネット空間に蓄積され続けている。そこで、インターネット空間のビッグデータを整理し、さらにコンピューター可読にすることを意図して、WWWの提唱者Tim Berners-Leeは、1999年に、世界中に分散するデータを分散型で疎につなげた知識のWeb (クモの巣)の構築を目指すセマンティックWebの構想を発表した。セマンティックWebとは、Webページに対するメタ情報(Resource Description Framework; RDF^{*18}などで定義)と意味情報(オントロジーで定義)を導入するものであり、このことにより、コンピューターが知識を活用する基盤が整備された。そして、その知識の源泉としてのデータの質と量が重要になってきた。

データの観点に基づくと、ディープラーニングに代表される機械学習技術では、大量の「学習データ」が必要である。インターネットやSNSで公開されているテキストや画像、動画に加え、IoT機器からの情報など、ある意味、ビッグデータを容易に入手できる環境は整いつつある。IDCの調査^{*19}によれば、2020年に生成されたデジタルデータは59ZB(ゼタバイト：1ZB = 2⁷⁰バイト = 10²¹バイト)を超えており、2025年には、187ZBのデータが1年間に生成されるという。

こうした大量のデータには、従来の定型データとは違う処理方法が必要になってくる。ビッグデータに対応する技術を「データ基盤技術」と呼ぶ。データ基盤技術には大きく三つがあり、一つは大量の

* 18 <https://www.w3.org/2001/sw/wiki/RDF>

* 19 “Worldwide Global Data Sphere Forecast, 2020-2024: The COVID-19 Data Bump and the Future of Data Growth” Apr 2020 - Market Forecast - Doc # US44797920

データを低コストでシェアするための「オープンデータ技術」である。オープンデータ技術には、セマンティックWebやLinked Open Data (LOD)^{*20}などがあり、これらに付随して「知識」の表現の共通化といった問題がある。残りの二つは、大量データを実用時間内に処理するための「データ処理基盤技術」、そして、データに含まれる個人に関わる情報を保護しつつ、データを有効に活用する「データ保護技術」である。AIとデータおよび知識の関わり方の歴史的発展については、「AI白書2019」の「2.4 知識処理とデータ」や「AI白書2020」の「2.2.7 知識」を参照されたい。

8 身体性

ロボティクスおよび「身体性」は知能にとっては重要な要素を含んでいる。物理的な身体の有無に関わらず、知的システムはすでに多く構築されており、身体性はAI設計において必要ないとの議論もある。しかし、それらのシステムの前提となる知識は、設計者側で準備されている場合が多く、設計者自身の身体性を通じて獲得された知識であるといえる。当然のことながら、それらを前提として有用なシステムを構築することは推奨されるであろうし、最近では、大量のデータや超高速大容量メモリを用いて、新たな知識発見の挑戦的研究もあり、AIは大きな成果を挙げている。

一方、人間と同じような知能のあり方を実現し、人間と共生する人工物を設計しようとするとき、現状の人工システムのさまざまな能力は、まだまだ劣っており、そのハンディキャップを考えると、物理的な身体を前提に考えることこそが重要であり、むしろ有利に働くとさえ考えられる。物理的な身体、すなわちTVカメラやマイクロフォンなどのさまざまなセンサーや身体を動かすアクチュエーターがあることで、ロボットが置かれた環境との多様な相互作用を考慮しなくてはならない。システム自身が、自分がどのような身体を持ち(身体表象)、どのような動きが可能なのか(運動の主体感、所有感覚)を知ることがもちろんのこと、その先には自己と他者の区別(自他認知)や、他者の動きや感じ方を推定(ミラー・ニューロンシステム^{*21}の働き)したりすることで、他者に対する思いやり(共感)などの行動すら生み出されることが期待されることになる。物理的な身体を通じてのロボットの能力を表現する方法においては、その事自体が受け入れる人間側にロボットの能力を推定可能(透過性)にすることを期待できるため我々の安心感にもつながることになる。AIが身体を手に入れることで初めて本当の安全安心が実現するのかもしれない。

ディープラーニングによりロボットの学習も大きく変化した。ルールベースで記述された動作ではなく、試行錯誤でその場に合わせた動作を学習することが可能になったのである。これらは深層強化学習の応用であり、自動運転でも試されている。さらにロボットが他者を含む環境と相互作用する際の、感覚や運動の予測の誤差を最小化する規範を認知発達の原因として扱ういくつかのアプローチなどは、AIでの予測と非常に親和性が高いものである。これまで述べたさまざまな知能がすべて身体を伴うロボットとして具現化するという考え方もあり、行動する知能と安心安全のための知能の実現として重要な分野であるともいえる。身体性としてのロボティクスのイメージを図表(付)11-7に示す。

ディープラーニングは、大量のデータを用いてデータに内在する特徴量を学習する。現在得られている結果からはその高い精度から「眼」を使った人間の仕事のすべてを機械化できる「可能性」に期待が

* 20 <https://www.w3.org/DesignIssues/LinkedData.html>

* 21 di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. Understanding motor events: a neurophysiological study. *Experimental brain research*, 91(1), 176-80. (1992)

集まっている。データによる学習により外界の認識が実現すれば、外界に働きかけた結果から自動的に「身体知」を学習する研究も飛躍を遂げ、ロボットや産業機械の研究がホットな分野となる。自然言語に基づいた自動的な「意味」処理も可能になり、自動的な知識獲得も可能となるであろう。

他にも柔らかな身体により環境との相互作用を豊かにするソフトロボティクスの研究分野やこれからのロボティクスの在り方については「AI白書2020」の「2.2.8 身体」を、人工物との共生社会への思想的背景については、「AI白書2019」の「2.5 身体性とロボティクス」を参照されたい。

図表(付)11-7 身体性としてのロボティクス



9 創造

ディープラーニング技術は、識別や回帰だけでなく、これまで存在していなかったデータ、たとえば架空の画像の生成にも利用できる。一般に「生成モデル」と呼ばれるモデルを使うことで、AI技術による「創作」が可能になる。画像の生成(描画)だけでなく、文章やテキストの要約など^{*22}、これまで人間が行ってきた行為に迫る「作品」が創られつつある。

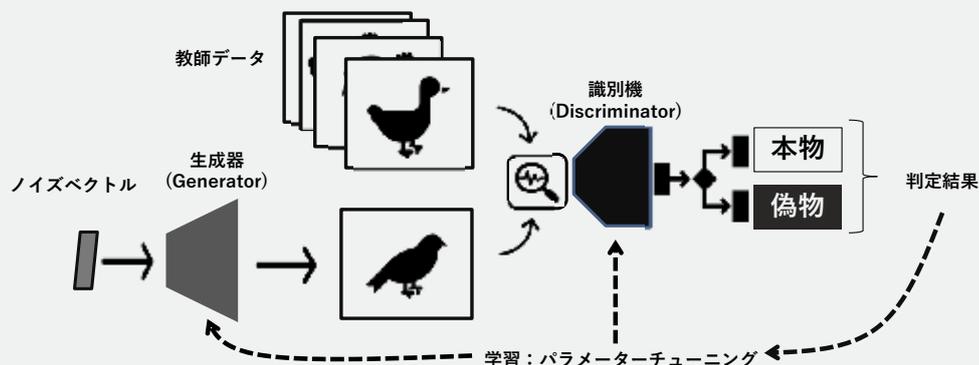
生成モデルとは、識別モデル(画像をみてネコと認識するようなモデル)と対になる概念で、ともに認識処理などに使われるものである。どちらのモデルもネコの画像を見せて「ネコ」と判定することが可能だが、生成モデルを使うと、ネコというクラスを指定してネコの画像を「生成」することが可能になる。モノクロの写真や映画をカラー化することも行われつつある^{*23}。すでに生成モデルを利用したディープラーニングを使い、過去のモノクロテレビ番組をカラー化して放送した例もある。深層ニューラルネットワークを使った生成モデルは深層生成モデルと呼ばれる無関係な二つの画像を使い、たとえばネコの写真とゴッホの画像からゴッホ風のネコの絵画を生成するといったことも可能になってきた。すでにインターネットでは、こうした深層生成モデルを使ったサービス^{*24}などが運用されている。

* 22 たとえば Adobe Sensei のテキスト自動要約
<https://business.adobe.com/jp/products/experience-manager/sites/auto-text-summarization.html>

* 23 <https://www.nhk.or.jp/str1/publica/rd/182/8.html>

* 24 <https://datagrid.co.jp/all/release/33/>

図表(付)11-8 GANの概要



代表的な深層生成技術には、オートエンコーダーを利用した「変分オートエンコーダー」(Variational Autoencoder; VAE)や「敵対的生成ネットワーク」(Generative Adversarial Networks; GAN)があり、医療の画像診断など向けのデータの生成や、不良品検知、商業デザインの作成、写真素材の作成、音声合成、白黒動画への自動着色、画像の超解像化などに応用されている(図表(付)11-8)。また自動運转向けの動画データの生成、動画の高解像度化、動画からのキャプション生成などへの応用や、さらには、テキストからアニメーションを生成するなどの応用も進められている。最近では、ノイズによる画像の破損過程を詳細に解析し、破損プロセスを逆転させることで機能するDiffusion Modelsが提案され画像、音声において最高性能を示している^{*25}。最近ではdeep fakeに代表されるように元画像に別の画像を重ね合わせることで、たとえば写真の中の人物を全く別の人物にすり替えてしまう技術も登場している。先端AI技術の倫理的側面が今後は重要になる。

* 25 High Fidelity Image Generation Using Diffusion Model
<https://ai.googleblog.com/2021/07/high-fidelity-image-generation-using.html>