

第 2 章 人工知能

浅川伸一

02/Feb/2019

1 はじめに

本章では人工知能について解説します。人工知能について最近ではほぼ毎日報道されるようになりました。マスメディアは騒ぎ過ぎのように見えますが、それでも次のような事柄は大きく取り上げられたようです。

1. 音声認識や画像認識で人間の認識能力を越えた (2015 年)
2. 囲碁やテレビゲームを解くプログラムが人間の世界チャンピオンを破った (2016 年)
3. 自動翻訳の性能が格段に向上した (2016 年)
4. フェイク画像, フェイクニュースを生成する人工知能が話題となった (2017 年)
5. フェイク動画が話題となった (2018 年)
6. 自然言語処理でも人間の成績を上回った (2018 年)

これらの報道を始めとして多くの関連する報道がなされています。加えて自動運転がもう直ぐ実現するか、人工知能が人間の仕事を奪うという人工知能脅威論まで聞こえてくるようになりました。この章の目的は巷(ちまた)に溢れる報道に惑わされることなく、どのようなことが起こっているのかを考える材料を提供することを目指しています。

ではなぜ、毎日のように人工知能関連の報道を見聞きするようになったのでしょうか。まず最初にこのことを考えてみましょう。

ある人、モノ、出来事、製品の出現によって、既存の考え方、慣習、規則、常識が一変してしまう劇的な変化が起ることを、英語ではゲームチェンジャー (game changer) と言ったりします。人工知能におけるゲームチェンジャーを 1 つだけあげるとすれば、畳込みニューラルネットワーク (Convolutional Neural Networks: 以下 CNN と略記します) です。本章第 4 節では CNN を簡単に説明してあります。

本章では次のような構成になっています。最初に知的であるとはどういうことかを改めて問い直すことから始めます (第 2 節)。続く第 3 節では人工知能の歴史を簡単にまとめています。第 4 節ではゲームチェンジャー CNN を詳しい解説を試みました。第 5 節では 3 つのニューラルネットワーク手法を示してあります。最後の 6 節では人工知能の意味を筆者なりに考えて結果を記してあります。

2 知能とは何か

2.1 人工知能研究

何をもって知的であるとするかは、知能の定義が定まっていない以上決めようがありません。知能検査で測定された数値が知能であるとする考え方もあります。ですが、その考え方では、知能検査で測ることができない知能は存在しないかのような錯覚を引き起こしてしまいます。このとき、もしある考えに基づいた知能が本当の知能を説明し尽くしているのであれば、その考えに沿って実際に作ってみることができるだろうという考え方があります。実際に作ってみることで、その考えが正しいかどうかを試すことができるからです。このような方法を構成論的アプローチと呼びます(谷口, 2014)。ある考えが間違っているのであれば正しく動作しないでしょうし、正しければ正しく動作するはずです。構成論的アプローチの良いところは、枝葉末節の細かい枝葉にこだわらず、全体が俯瞰(ふかん)できることです。断片的な知識を積み重ねても全体像が見えないことがあります。これに対して構成論的アプローチは動作する全体像を考える必要があるので全体像を見失うことが少ないと言えるでしょう。一方で、未だよく分かっていない部分は類推して作る必要があるので、全体として動作するにしても細部は実体とかけ離れている場合もあります。構成論的アプローチの対極にある方法を分析的アプローチと呼び心理実験、生理学的実験の多くがこのアプローチに該当します。どちらのアプローチがだけが正しいというわけではなくどちらのアプローチも必要なのだと理解して良いでしょう。従って人工知能の研究とは知的な振る舞いを実際に作って実証してみるという意味で知能の構成論的アプローチであり、人間知能を扱う心理学で明らかとなった知見をコンピュータやロボットの上にも実現する研究でもあります。この意味では人工知能と知能の心理学とは重複する領域であるということが出来ます。

人工知能の反対の概念を自然知能(natural intelligence)と呼びます。人間を含めて生物の持つ知的な振る舞いを自然知能と呼び、自然知能を模倣しようとする試みを人工知能と呼びます。人工知能を実現するためにはコンピュータが必要です。ですから人工知能研究の歴史は最長でもコンピュータの歴史と同じこととなります。以下に、コンピュータを人間のような振る舞いさせることを考えた最初の研究者であるアラン・チューリング(Alan Turing)の考えたことを紹介します。

2.2 知的であるとはどういうことか-チューリングテスト-

ある機械が知的な振る舞いをすることを測る方法に、チューリングテストがあります(図1)。アラン・チューリングは「機械は考えることができるか?」という問いを考えました(Turing, 1936)。

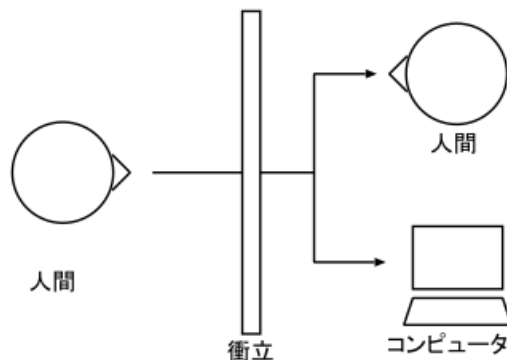


図1 チューリングテストの模式図

60年以上前にチューリングが考えたチューリングテストは、現在でも形を変えて用いられることがあります。人工知能の歴史は、チューリングテストに合格するコンピュータプログラムを作る努力であるとも言えます。チューリングテストを人工知能最古の歴史だとすると、人間の脳を模倣する試みや、脳の働きを数式として表現する研究、人間の替わりをする江戸時代のからくり人形を含めたロボットを考えれば、これの方が人工知能の研究より歴史が古いと言えます。人間の脳、ロボット、人工知能の3者を結びつけて考える研究者もいます。ロボットが自立して動くためには周囲の状況を正しく認識する必要がありますが、そのためには物体認識や音声認識などをロボットが行える必要があるからです。物体認識や音声認識は人工知能の一分野とも考えられますが、そのためには人間の物体認識や音声認識の仕組みを知る必要があります。このように人間の脳、ロボット、人工知能は関連します。本書は知能の本ですので、ロボットの話題は軽く触れるだけ止めます。

3 人工知能の歴史

本節では、人工知能の歴史を概説します。人工知能を大きく区分すると、ニューラルネットワーク系と記号処理系になりますが、現在までにそれぞれ3回の流行がありました。人工知能研究の歴史を理解するにはこの3回の流行でどのようなことが起こったのかを知る必要があります。

人工知能の教科書を紐解く(ひもとく)と、人工知能という言葉が初めて用いられたのは、1956年にアメリカのニューハンプシャー州にあるダートマス大学で開催されたダートマス会議であると書かれています。ですがそれ以前にもチューリングの研究はありましたし、日本のからくり人形まで含めれば、機械や人形に知的なふるまいをさせる試みはずっと以前から存在しました。

ダートマス会議では人工知能が取り組むべき7つの問題を列挙しています*1。

1. コンピュータの自動化 (Automatic Computers)
2. 日常言語を用いたコンピュータプログラミング (How Can a Computer be Programmed to Use a Language)
3. ニューラルネットワーク (Neuron Nets)
4. 計算規模の理論 (Theory of the Size of a Calculation)
5. 自己改善 (Self-Improvement)

*1 [A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence](#)

6. 抽象化 (Abstractions)

7. 乱雑さと創造性 (Randomness and Creativity)

上記の解くべき問題のリストは、あらかじめ定められた課題だけしか扱うことができない融通の効かない機械を越えて人間のように柔軟に状況に対応することが人工知能に求められた課題であると言えます。

人工知能研究には 2 つの大きな流れがあります。一つは脳がニューロンを基本単位としたネットワークであることから、ニューロンのネットワークを模倣しようとする試みでニューラルネットワーク研究です。もう一つは、他の種と異なり我々人間だけが、言語や数学など高度な記号処理体系を持っていることから、これらの記号処理体系をコンピュータに実現しようとする記号処理に基づく人工知能研究です。

3.1 第一次ブーム

記号処理系の人工知能ブームは先のダートマス会議が始点となります。一方で、ニューラルネットワーク系の研究は、パーセプトロン (perceptron) が出発点となりました (Rosenblatt, 1958)。パーセプトロンとは「知覚するもの」あるいは「知覚する装置」といった意味の言葉です。図 2 にパーセプトロンの概略を示しました。

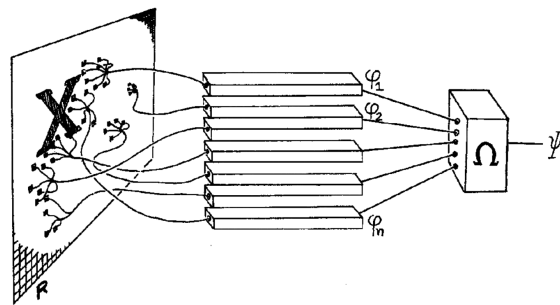


図 2 パーセプトロンの概念図 (Minsky & Papert, 1988) より

図 2 の左側にある入力画像からニューロンの連絡を介して右側へと信号が伝達されます。脳の神経細胞が電気的な信号のやり取りをしています。一つ一つのニューロンは活動するか、または活動しないかのどちらかです。パーセプトロンはこのようなニューロンの活動を模倣した人工ニューラルネットワークでした。図中の $\varphi_1, \varphi_2, \varphi_n$ がニューロン間の結合の強さを表しています。ニューロン間の結合の強さが変化することによって、経験を通して入力画像を認識するようになります。ニューロンの結合が左から右へと一方向だけしかありません。このようなニューラルネットワークをフィードフォワード型のニューラルネットワークと呼びます。パーセプトロンは経験を通して学習する能力を持っています。パーセプトロンの学習とは φ の値が変化することを指します。パーセプトロンは経験を通じて学習することができますが、先述の通りパーセプトロンの学習とは図 2 の φ の部分の変化を指します。学習する部分が φ 一箇所であるので学習できる内容、すなわち画像の識別能力には限界がありました。この学習能力の限界のために第一次ブームは下火になります。

一方の記号処理系の人工知能ブームはサー・アーサー・サミュエル (Sir Arthur Samuel) によるチェッカーと呼ばれるボードゲーム (図 3) を解く人工知能プログラムが挙げられます。

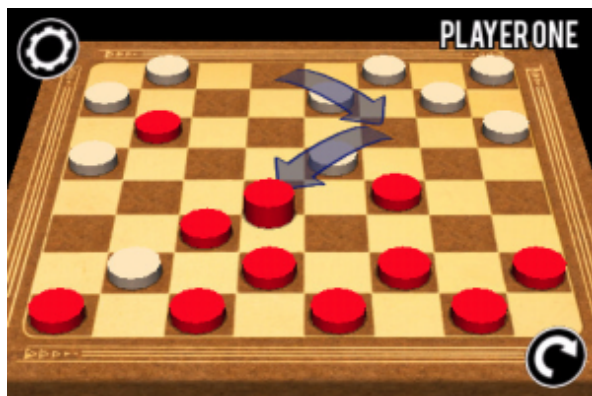


図3 チェッカー

チェッカーのようなボードゲームは自分の手番で選択可能な手に対して、それに応じた対戦相手の選択の可能な手が決まります。このように考えれば、ゲーム開始時点の時初期状態として、交互に駒を動かし行くこととなります。自駒の動きが複数手選択可能ならば、どの分岐を選択するかが勝負を左右することとなります。相手も同様に分岐を選択することとなりますので、ゲーム全体を、可能な選択肢を上から下へ向かう選択肢の木のように表現することができます。

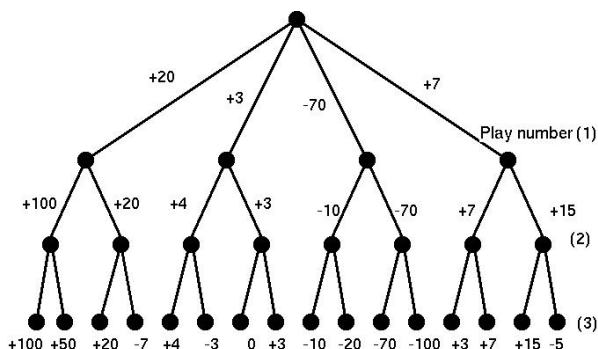


図4 ゲーム探索木

ボードゲームは図4のように木を探索することと同じ意味です。木で表された選択肢のうち、どの選択肢を選ぶと勝ちにつながるのかがその局面での評価になります。図では正負の値によって、評価の良し悪しが表現されています。その都度の局面で、最終的に最大の値を得ることが可能な選択肢を選ぶことが勝利につながります。

木が深くなると探索が大変になります。従ってチェッカーよりも選択可能な手数が多いチェス、将棋、囲碁などは当時のコンピュータでは解くことができませんでした。

3.2 第二次ブーム

第2次ブームは、記号処理、ニューラルネットワークとともに1980年代になります。

ニューラルネットワークについては、パーセプトロンの学習能力の限界を乗り越える手法が提案されたことからブームが始まります。パーセプトロンはニューロン間の結合を遡って更新(学習)することができなかつ

たので、単層のニューラルネットワークでした。誤差逆伝播法 (バックプロパゲーション) と呼ばれる方法が提案されたことにより、パーセプトロンが持っていた能力限界を越えることが可能となりました (Rumelhart, Hinton, & Williams, 1986)。

記号処理系の人工知能はこの時代、専門家の知識を表現したエキスパートシステムと呼ばれるシステムが代表的です。この時代までのコンピュータの処理能力の進歩に伴って、専門家の知識をルールの集まりとして表現して専門家の知識を肩代わりしようとする試みがなされました。またコンピュータの処理能力の進歩は、チェッカーよりも探索木が大きくなるチェスでも世界チャンピオンに肩を並べるまでになりました。

ところが、エキスパートシステムが専門家の知識を引き出すために労力を要すること、誤差逆伝播法よりも数学的に洗練されているサポートベクターマシンが提案されたこと (Vapnik, 1995)、などの理由によりブームは去り、二度目の冬を迎えました。

ニューラルネットワーク、記号処理の両者ともコンピュータの処理能力の進歩の恩恵を受けたと言えます。ですが、今日の我々から見れば、コンピュータの処理能力はこの時代の研究の限界も示していたことになりません。この時代のコンピュータの能力は今日のスマートフォンの処理能力に劣ります。そのようなコンピュータの処理能力のため、応用範囲が限定されていたことが指摘できます。

3.3 第三次ブーム

21 世紀に入ると記号処理系の人工知能研究は確率的、統計的な処理を視野に入れ発展してきました。最も大きな事件はコンピュータ処理能力向上と処理容量拡大により大規模で実用的な問題を解くことが可能になってきたことが挙げられます。大規模で実用的な問題を解くための工夫が提案され蓄積されてきましたが、2009 年から始まった大規模画像認識コンテスト (通称イメージネット) で 2012 年になって本章冒頭で紹介した CNN を用いたカナダ・トロント大学のチームが、サポートベクトルマシンを用いた認識性能を 10 パーセント以上引き離して優勝したことで耳目を集めました。

図 5 は大規模画像認識コンテスト (通称イメージネットコンテスト ILSVRC) の成績を示しています。

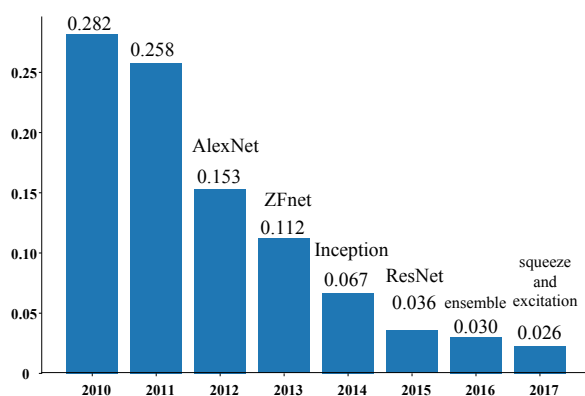


図 5 イメージネットコンテスト優勝チームの成績の変化

横軸は開催年度で、縦軸はその年の優勝チームの成績を表しています。およそ 130 万枚の画像を 1000 種類のカテゴリに分類するコンテストですが、その成績はモデルが予測するカテゴリの上位 5 つの中に正解が

含まれているか否かで競われます。図は間違えた割合を示してありますので、棒グラフが短い、あるいは小さいほど成績が良かったことを表しています。130 万枚の画像の中には紛らわしい画像が含まれていますので、たとえ人間であっても 100 % 正解はできません。人間の場合 5 % 強は間違えるようです。ところが 2015 年の優勝チーム（マイクロソフトアジア研究所のチーム）は人間の誤り率よりも低い値を叩き出しました (He, Zhang, Ren, & Sun, 2015)。2016 年に入って他のモデルも人間を超える性能が得られたと報告されています。すなわち大規模画像認識コンテストの成績を見る限りコンピュータの認識性能は人間を越えたと言って良いでしょう。人間の成績を超えたことは世界的な衝撃を与えました。

技術的には、中間層を多層化する工夫が実用化されてきたため認識精度が向上しました。第一次ブームのパーセプトロンでは、学習が最上位のニューロンとその直下のニューロンとの間の結合だけに限定されていたのに対して、第二次ブームの誤差逆伝播法では更にその下位に位置するニューロン間の結合についても学習することができました。第三次ブームのニューラルネットワークは更に深い層のニューロン間の結合についても効率よく学習する工夫がなされていました。

CNN の特徴として、次の 7 を上げることができます (浅川, 2016b)[^2017Aakawa_deep_jps]。

1. 非線形活性化関数 (nonlinear activation functions)
2. 畳込み演算 (convolutional operation)
3. プーリング処理 (pooling)
4. データ拡張 (data augmentation)
5. バッチ正規化 (batch normalization)
6. ショートカット (shortcut)
7. GPU の使用

上記 7 つの特徴を説明するのは専門的になりすぎるので省略します。一つだけ説明するとすれば最後の GPU とは高解像度でしかも処理速度を必要とするパソコンゲームで用いられるグラフィックボードのことで、詳細な画像を高速に画面に表示する必要から開発されたグラフィックボードですが、大規模なニューラルネットワークの計算でも用いられる数学が同じです。そのため、ゲーム用に開発されたグラフィックボードがニューラルネットワークにも用いられるようになりました。

図 6 は画像認識の性能向上の例を示しています (Szegedy et al., 2015)。



図 6 CNN はエスキモー犬とシベリアンハスキーとを区別できる

画像認識だけでなく、音声認識、記号処理でも深層化したニューラルネットワークが用いられ性能の向上が報告されています。その一例が ATARI のゲームを解く DQN と呼ばれるニューラルネットワークの一種です (Mnih et al., 2015)。DQN は複数のゲームを解く際に人間が個別にルールを与えて学習させるものではありません。CNN による画像認識能力を活用しているため、画面の入力は同じです。強化学習と呼ばれる手法で高

得点を得るためのボタンやジョイスティックの動きを学習することでゲームによっては人間を上回る成績を示しました。図 7 に結果を示しました。図では一行に一つのゲームの成績が示されています。もっとも上達した最上行のピンボールでは人間の成績の 2359 パーセント、2 つ下のブロック崩し (breakout) でも人間の成績の 1327 パーセントとなっています。

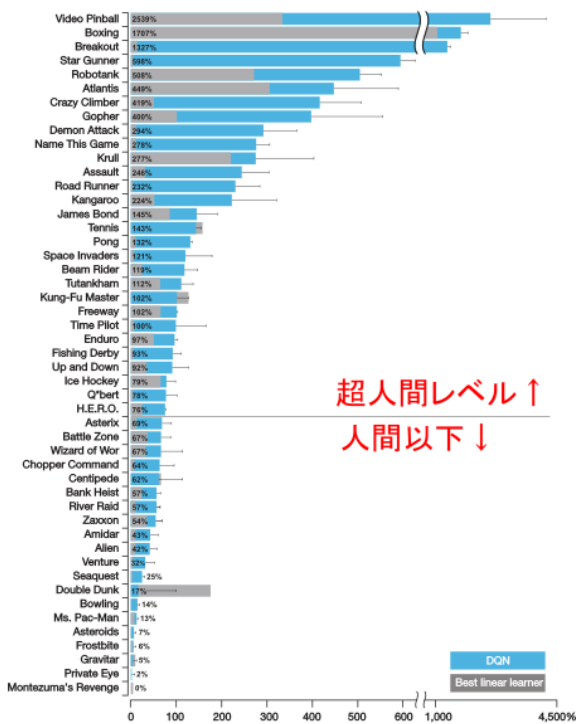


図 7 DQN の結果 Nature より

DQN のノウハウを活かしてチェスよりも選択枝数が多い囲碁に挑んだのが、グーグル傘下のスタートアップチームであるディープマインド社が開発したアルファ碁でした。

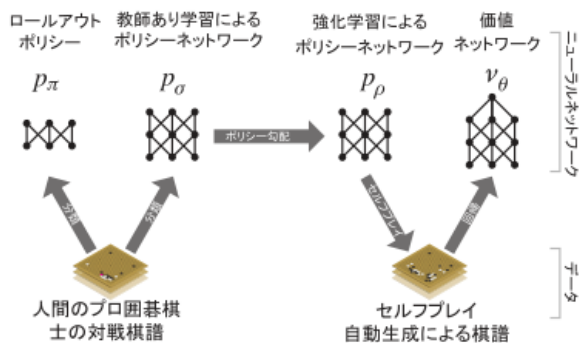


図 8 アルファ碁 Nature より Silver et al. (2016)

アルファ碁は囲碁の世界チャンピオンを破っただけでなくその後も進歩を続けています。囲碁のプロの対戦棋譜から学習していますが、棋譜の足りないところは自らのセルフプレイによって学習します。その結果、囲

碁の定石が覆ることさえ起こっているようです。

同時期にグーグルの機械翻訳の精度が向上し話題となりました (Wu et al., 2016)。このことは記号処理系の人工知能にもニューラルネットワークの成果が影響を及ぼしていることを示しています。第三次ブームは、記号処理系の成果以上にニューラルネットワークの性能向上が貢献していると言えます。

4 畳込みニューラルネット (CNN) とは何か

本節では深層学習、特に CNN と呼ばれるニューラルネットワークについて解説します。

最初に画像処理の概略を述べる CNN が、それまで主流であった従来の手法の性能を凌駕したことはすでに述べました。CNN の特徴の一つに エンドツーエンドと呼ばれる考え方があります。エンドツーエンドとは、従来手法によるパターン認識システムでは、専門家による手の込んだ詳細な作り込みを必要としていたことと異なり、面倒な作り込みをせずとも性能が向上したことを指します。

エンドツーエンドなニューラルネットワークにより、次のことが実現しました。

- ニューラルネットワークの層ごとに、特徴抽出が行われ、抽出された特徴がより高次の層へと伝達される
- ニューラルネットワークの各層では、比較的単純な特徴から次第に複雑な特徴へと段階的に変化する
- 高次層にみられる特徴は低次層の特徴より大域的、普遍的である
- 高次層のニューロンは、低次層で抽出された特徴を共有している

このことを簡単に説明してみます。

我々人間は、外界を認識するために必要な計算を、生物種としての発生の過程と、個人の発達を通しての経験に基づく認識システムを保持していると見ることができます。従って我々の視覚認識には化石時代に始まる光の受容器としての眼の進化の歴史と発達を通じた個人の視覚経験が反映された結果でもあります。人工知能の目標は、この複雑な特徴検出過程をどうやったらコンピュータが獲得できるかということでもあります。外界を認識するために今日まで考案されてきたモデル (例えば、ニューラルネットワークサポートベクターマシンなどは) は複雑です。ですがモデルを訓練するための学習方法はそれほど難しくありません。この意味で画像認識課題が正しく動作するためのポイントは、認識システムが問題を解く事が可能なほど複雑であるかどうかではなく、十分に複雑が視覚環境、すなわち画像認識の場合、外部の艦橋を反映するために十分な量の像データを容易にすることができるか否かにあります。今日の CNN による画像認識性能の向上は、簡単な計算方法を用いて複雑な外部環境に適応できる認識システムを構築する方法が確立したからであると言うことが可能です。

図 9 に画像処理の例を挙げました。図 9 では入力画像がネコであるか否かを判断する画像認識であるとなりました。我々はネコの画像を瞬時に判断できます。ですが画像認識の難しさは、入力画像が図 9 に示されているように入力信号の数字の集まりでしか無いことです。このようなデータを何度も経験することでネコを識別できるようにする必要があります。

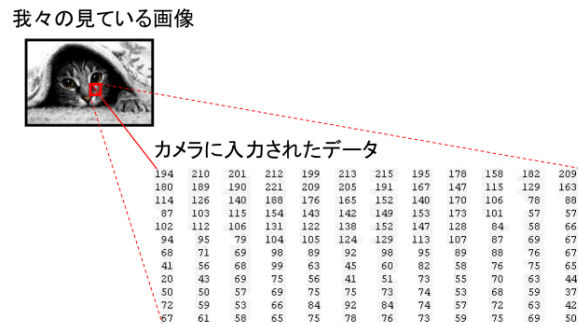


図9 我々の見ている画像は数値の列としてデータ化される

図9に示したようにコンピュータに入力される画像は数字の塊に過ぎません。

状況ごとにとるべき操作を命令として逐一コンピュータに与える指示する手順の集まりのことをコンピュータプログラムと呼びます。人間がコンピュータに与えることができる操作や命令によって画像認識システムを作る場合、命令そのものが膨大になったり、そもそも説明することが難しかったりします。例を挙げれば、お母さんを思い浮かべてくださいと言われれば誰でも、それぞれ異なるイメージであれ思い浮かべることができます。また、提示された画像が自分の母親のものであるか、別の女性であるかの判断は人間であれば簡単です。ところがコンピュータには難しい課題となります。加えて母親の特徴をコンピュータに理解できる命令としてプログラムすることも難しい課題です。つまり自分の母親の特徴を曖昧な言葉でなく明確に説明するとなるととても難しい課題となります。というのは、女性の顔写真であればどの写真も似ていると言えるからです。顔の造形や輪郭、髪的位置などはどの画像も類似していることでしょう。ところがコンピュータにはこの似ている、似ていないの区別が難しいのです。

加えて、同一ネコの画像であっても、被写体の向き視線の方向や光源の位置や撮影条件が異なれば画像としては異なります。図10に示したように入力画像の中の特定の値だけを調べてみても、入力画像がネコである、そうではないかを判断することは難しい課題になります。

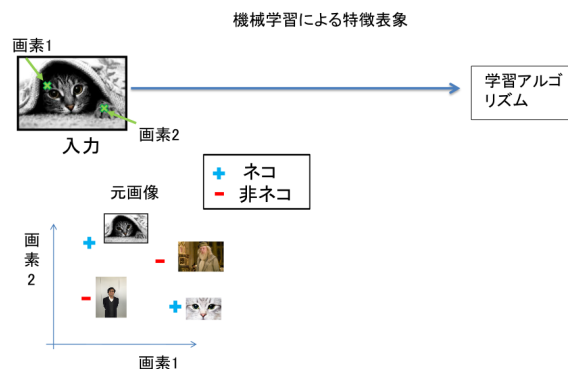


図10 数値列から画像を理解することは難しい

現在の画像認識では、特定の画素の情報に依存せずに、入力画像が持っている特徴をとらえるように設計されます。たとえば、ネコを認識するために必要ことは、ネコに特徴的な「ネコ目」や「ネコ足」を検出するこ

とであると考えます。入力画像から、ネコの持つ特徴を抽出することができれば、それらの特徴を持っている入力画像はネコであると判断して良いことになります (図 11)。

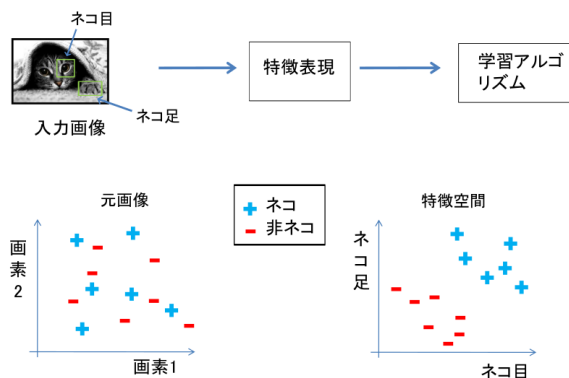


図 11 適切な特徴抽出ができればネコ目特徴とネコ足特徴が同時に高い値となる画像はネコと認識して良い可能性が高まる

図 12 は、音声認識と画像認識の両分野において CNN が用いられる以前の従来手法をまとめたものです。

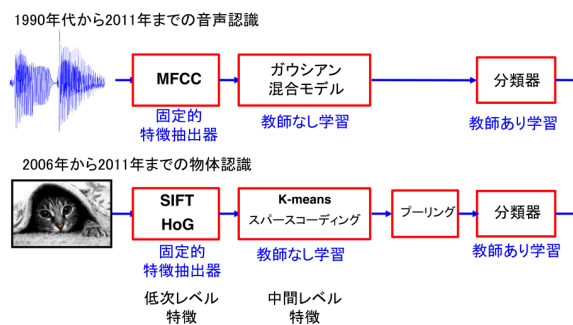


図 12 従来主流であったパターン認識システムの構成

図 12 のような従来手法に対して、CNN ではエンドツーエンドな特徴抽出を多層多段に重ねることによって複雑な特徴を抽出 (検出) しようとしています (図 13)。

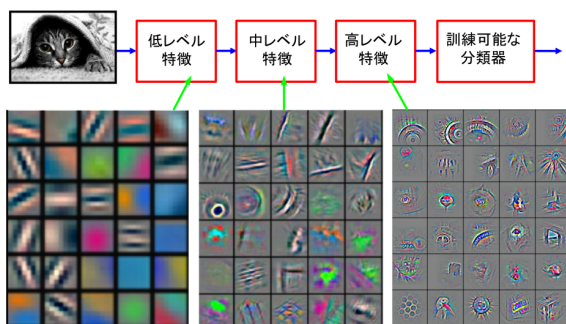


図 13 非線形特徴変換を多数回繰り返した学習器を深層学習 (ディープラーニング) という

コンピュータにはネコ目特徴検出器，ネコ足特徴検出器は備わっていません。そこで画像認識研究では，画像の統計的性質に基づいて特徴検出器を算出する方法を探す努力が行われてきました。しかし，コンピュータにネコ目特徴やネコ足特徴を教えるは容易なことではありません。このことは画像処理の分野だけに限りませんが，音声認識でも言語情報処理でもそれぞれの特徴器を一つ一つ定義し，チューニングするのは時間がかかり，専門的な知識も必要で困難な作業でした。

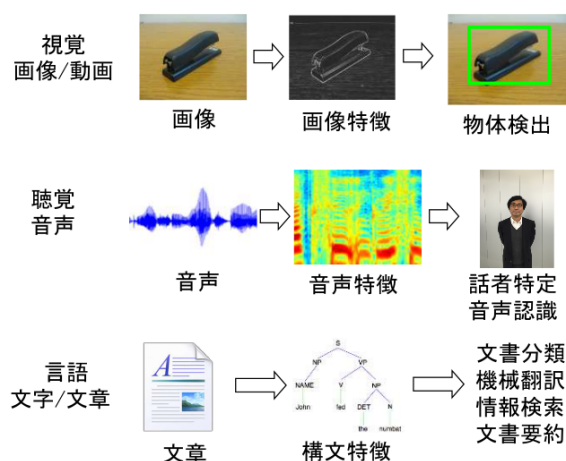


図 14 コンピュータはどうやって知覚するか。画像 (上)，音声 (中)，言語 (下) とも入力情報 (左) から特徴抽出 (中) して分類器 (右) で認識に至る

まとめると，1950 年代後半以来：固定的，手工芸的特徴抽出器と学習可能な分類器を用いた認識システムを作ることが試みられてきたといえます。これに対して CNN が主流となった現在はエンドツーエンドで学習可能な特徴抽出器を多数重ね合わせることで性能が向上しました。

夢のような話が続きましたが，本節の最後に逆に CNN は簡単に騙すことができる例を挙げておきます (図 15)。図では，左の画像が入力画像で，CNN は確信度 57.7 パーセントでパンダである認識しました。ところがこの画像に 0.007 だけ意味のない画像 (図中央) を加えた画像 (図右) を CNN は 99.3 パーセントの確信度でテナガザル (gibbon) と判断しました。この例はここでは詳しく触れることはしませんが 敵対的学習と呼ぶ訓練手法を説明する際に用いられた例です (Goodfellow, Shlens, & Szegedy, 2014)。

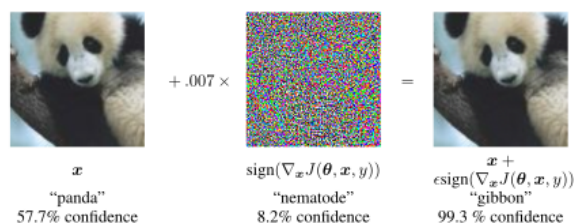


図 15 パンダの画像に 0.7% の雑音を乗せるとテナガザルと認識してしまう例

この例からも分かることは以下のようにまとめられるでしょう。すなわち，人間の脳を模したニューラルネットワークである CNN が大規模化画像認識チャレンジにおいて人間の認識性能を越えたと報道されました。ですが，人間の視覚認識を完全に実現したと考えるのは早計で，解くべき課題は未だ多数あるということ

す。この状況は、音声認識や言語情報処理でも同様であると言えます。

5 人工知能の現状

前節では CNN の多層化と表現について説明しました。本節では CNN に留まらず広く最近の人工知能研究の傾向をとらえることを試みます。← - 最近の人工知能研究の特徴 - 最近の人工知能の趨勢を理解する上でのキーワード → 用いられているアルゴリズムから分類すると次の 3 つが代表的です。

1. 畳み込みニューラルネットワーク (CNN)
2. リカレントニューラルネットワーク (RNN)
3. 強化学習 (RL)

以下では簡単にそれぞれを説明してみました。

5.1 畳み込みニューラルネットワーク (CNN)

畳み込みニューラルネットワーク CNN は近年、画像認識や音声認識で急激な性能向上をもたらしました。ニューラルネットワークとは、人間の脳の振る舞い（神経回路）を模した計算モデルを指します。現在の第三次人工知能ブームの火付け役となったのは、深層学習（ディープラーニング）という機械学習の手法です。ここで用いられているモデルがニューラルネットワークであり、ニューラルネットワークの中間層を多層化したモデルのことを深層学習と言ったりします（浅川, 2016a）。

5.2 リカレントニューラルネットワーク (RNN)

リカレントニューラルネットワーク (RNN) とは、時間的な変化や順序といった系列情報を扱うニューラルネットワークモデルです。このため音声認識、自然言語処理、ロボットの生成制御などに用いられています。時々刻々変化するデータを扱うには、それまでに処理されたデータの系列を文脈として保持しておく必要があります。リカレントニューラルネットワークを拡張した長短期記憶モデル (LSTM: Long short-term memory) を用いる場合が多いです。

応用事例としては、最近マイクロソフトが開発した女子高生人工知能「りんな」の対話生成アルゴリズムが有名です。LINE であたかも女子高生と会話しているかのようなコミュニケーションができることで話題になりました。

そのほかにも自動翻訳、画像と文章と相互変換（画像を入力するとその画像を説明する文章を生成する、逆にある文章を与えると対応する画像を生成する）などがあります。リカレントニューラルネットワークを用いた技術は他の従来手法の性能を上回り、現時点での最高性能と認められています。

現段階で、人間が書いたものなのかコンピュータが書いた文章なのか区別がつかない場合すらあります。新たなチューリングテスト（あるいはチューリングチャレンジ、コンピュータの生成した文か人間が書いた文か見分けがつかなければ、もはやコンピュータに知性が宿っていると言っても良いだろうという考え方）と言ったりもします。

5.3 強化学習 (RL)

三つ目は 強化学習 です。強化学習という言葉は古い言葉ですが機械学習の文脈では、環境とその環境におかれた動作主（エージェントと言ったり、ロボットシステムだったりします）が、環境と相互作用しながらより良い行動を形成するためのモデルです。動作主は、環境から受け取った現在の状態を分析して、次にとるべき行動を選択します。このとき将来に渡って報酬が最大となるような行動を学習する手法の一つです。

2015 年には、Google 傘下のデープマインドというスタートアップチームが開発した囲碁プログラム AlphaGo がプロ棋士のイ・セドル氏に勝利し話題になりました。AlphaGo は強化学習を基本技術の一つとして用いています。

6 人工知能の影響

ここまでは、人工知能研究についての話題を説明してきました。本章の最後となる本節では、人工知能にまつわる社会的、経済的な話題に触れておきます。

6.1 人工知能が仕事を奪う？

人工知能の問題でよく取り上げられる問題の一つに「トロッコ問題」があります。将来、汎用人工知能が登場して生き残れる産業とそうでない産業が生まれてしまうと仮定します。どちらかの産業を活かして、他方は捨てるあるいは、犠牲にしなければいけないジレンマ生じるのという問題です。

もう一つ、技術的な部分として「人工知能が望ましくない判断をした時に止められるのか？」という問題もあります。ある意識調査で自動運転車に緊急停止ボタン（キルスイッチ）は必要か？とアンケートをとったところ、必要だという方がほとんどだったそうです。実際には、Google の自動運転技術は人間が運転するよりも安全になっています。ですが、いざという時に人間が制御できないというのは怖いからです。

Google の発表によると、人工知能はすでにキルスイッチの無効化を学習できます。そのため、人工知能の暴走を防ぐには臨時割込判断を導入する必要があると言われています。

人工知能が反抗しないように、かつ臨時割込判断による仕事の効率低下などの負の効果を最小化するよう計画する必要があります。しかし、もし人工知能がこの計画を知っていたら、当然、その裏をかくように振る舞うと予想できます。人工知能が臨時割込判断の時だけ人間をだますようになると、どうすれば良いのか。これから考えていかななくてはならない技術的課題です。

これに関連しますが、哲学者ニック・ボストロムは著書『スーパーインテリジェンス』の中で次のように書いています。

現在のゴリラの運命は、ゴリラ自身以上に人類に依存している。我々人類の運命もいずれ機械（超知能）に依存するようになるだろう。つまり、ゴリラの運命を人間が操っていることにゴリラ自身が気づいていないように、超知能も我々に気づかせないようにする

6.2 人工知能とこれからの仕事

人工知能の利点とリスク*2 と題する興味深い論考が日本語に訳されました。平易な訳になっていますので目を通してみると良いでしょう。人工知能がもたらすであろう将来についての恐れる必要なければ、楽観視する必要もないことが書かれています。ですが残念なことに、この日本語訳には十分に根拠が示されていない箇所があります。必要な場合には、そのような主張の根拠は、元となる背景、議論を追いかけることができなければ、いかに尤もらしい主張であっても、単なる扇動に過ぎません。この日本語訳は主張の根拠や証拠を追いかけることができないという意味で残念な内容ですが、原文の英語の方には根拠、証拠、参考となる情報を追いかけることができるように配慮されています。重要な情報とは、このように必要であれば読者がその根拠は背景となるデータを追いかけることができる必要があるのですが、残念ながら多くのメディア（インターネットメディアに限らず、マスメディアも同様です）は、その根拠を明示することなく曖昧な情報を流しています。人工知能に関する情報も、全く同様で、不安を煽るような記事やニュース、乗り遅れると大変だとか、機械が仕事を奪うというような衝撃的な内容を掲載しています。最低限の判断基準として、根拠が提示されているか、論説の証拠や根拠を追いかけるための情報が示されていないニュースやサイトの情報は鵜呑みにしないという態度が必要でしょう。悪質なまとめサイトや流言蜚語に惑わされることなく、人工知能を正しく理解し、正しく活用できるようになってこそ我々の生活は豊かになることでしょう。

人工知能は第4次産業革命とも呼ばれます。それ以前の産業革命は、農村から都市部への人口流入、および1次、2次、3次産業に従事する人口構成の比率が変化したという社会構造の変化を伴いました。本章で示した人工知能技術により、かつての産業革命に伴って起こった社会構造の変化に比肩する変化を引き起こす可能性があるでしょう。たとえば、本章の冒頭でも述べた人工知能脅威論は、人工知能が人間の職を奪うという恐れからです。しかし、いわゆる3Kの（きつい、きたない、危険な）職業はむしろ人工知能を含む技術によって取って代わられた方が、人間らしい生活を送れることができるとも言えるでしょう。自動運転についても同様の議論が当てはまると考えられます。人間による自動車の運転には、人間特有の限界があるからです。人間の運転では、どうしても注意が一点に集中してしまう傾向があるために、児童の飛び出しなどの突発的で緊急の事態に対処しきれない状況が想定できるからです。

人間の注意の焦点が一点に集中する傾向があることを、人工知能はあらかじめ予想し、危険を回避することが可能になるとも予想されます。そうすると、むしろ自動運転の方が安全で好ましいとさえ言えるでしょう。このときに社会全体として考えなければならないことは、それでも交通事故が発生した場合に、その責任を誰の責めに帰すのかという、法律的、社会的、あるいは道徳的、倫理的問題になります。このような問題は人工知能研究者だけで解決する問題ではなく、社会を構成する全ての人々が議論して解決しなければならない問題であるとも考えます。

社会構造の変化は、現在と異なる新しい職業の創り出す可能性もあり、今現在我々が持っている常識や価値観が変化するものと予想されます。このように人工知能の進歩により、職業の選択や仕事の価値観も大きな変化が予想されます。我々の持つ知能にはこのような人工知能の進歩に対処するための柔軟さ求められることになるだろうと考えます。

*2 <http://futureoflife.org/background/benefits-risks-artificial-intelligence-japanese/>

引用文献

- 浅川 伸一. (2016a). 『Python で体験する深層学習』. 東京: コロナ社.
- 浅川 伸一. (2016b). 深層学習をめぐる最近の熱狂. 『基礎心理学研究』, 35, 149–162.
- Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv:1512.033835*.
- Minsky, M., & Papert, S. (1988). *Perceptrons* (Expanded Edition ed.; 中野 馨・坂口 豊, Trans.). Cambridge, MA: MIT Press. (パーセプトロン, 1993, パーソナルメディア)
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstr, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518, 529–533.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386–408. (In J.A. Anderson and E. Rosenfeld (Eds.) *Neurocomputing* (1988), MIT Press)
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructures of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.
- Silver, D., Huang, A., Arthur Guez, C. J. M. an, Sifre, L., Driessche, G. van den, Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, 529, 484–492.
- Szegedy, C., Liu, W., Jia, Y., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA.
- 谷口 忠大. (2014). 『イラストで学ぶ人工知能概論』. 講談社.
- Turing, A. M. (1936). On computable numbers, with an application to the entscheidungsproblem. *Proceedings of London Mathematics Society (Series 2)*, 42, 230–265.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser Łukasz, Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., & Dean, J. (2016). Google’s neural machine translation system: Bridging the gap between human and machine translation. 『arXiv preprint arXiv:1609.08144』.