

第4講 人間の脳の謎と深層学習の魔法－目を

持ったコンピュータが見せる未知の領域

【学習到達目標】

- ・ 3つの機械学習(教師あり/教師なし/強化学習)を説明できる。
- ・ 畳み込みニューラルネットワーク(CNN)について事例を挙げて説明できる。
- ・ カレントニューラルネットワーク (RNN)について事例を挙げて説明できる。

1. 第三次AIブーム「機械学習、深層学習の時代」

歴史的に見ると、1990年代に、情報検索、「データマイニング」、検索エンジンの3つの動きがありました。1992年、1) 米国政府主導によるTRC国際会議が開催され、情報検索と言う概念が登場しました。2) 必要なデータを引き出す「データマイニング」が国際会議やジャーナルで立上りました。3) 1998年には、検索エンジンのGoogle社が出現しました。2000年代に入ると、Web(ウェブ)が広がり、1) 1995年には1万サイトだったものが、2006年には1億サイトになりました。2) 「ビッグデータ」(大規模データ)という言葉が登場し、3) 大量のデータを用いた機械学習が実用化しました。2006年から、第三次AIブームに突入し、機械学習・深層学習の時代になりました。

・ 第3次AIブーム時代を拓いたジェフリー・ヒントン教授

「ディープラーニングの父」と称される「ジェフリー・ヒントン」教授は、英国生まれのコンピュータ科学者で認知心理学研究者です。ヒントン教授は「高校生のときに優秀な友人がいて、脳の機能についてよく話していました。大学に入り生理学を専攻して、脳の働きをさらに学べるのではないかと期待しましたが、講義ではそれを学ばませんでした。」と、インタビューで動機を語りました。ヒントン教授は、心理学やAIの講義を受けても求めていたものが得られなかったヒントン教授は「ニューラルネットワークを研究してみようと思立ちました。ニューラルネットワークの研究は、簡略化された脳機能の研究でもあります。」

と、ニューラルネットワークに関心を抱いたきっかけを話しています。

ヒントン教授のインタビューのポイントは「ニューラルネットワークの研究は簡略された脳機能の研究でもあります。ニューラルネットワークではプログラムは学習アルゴリズムをプログラムします。これはとても簡単です。ディープラーニングは、ニューラルネットワークが、層になったニューロンから知識を得るもの」です。

ヒントン教授はディープラーニングを可能にしたものは？と問われ、「ディープラーニングの画期的な点は、AIに知識を与えようとする時、知識自体をプログラムをする必要性がない点です。最も大きな成功を収めたアルゴリズムは誤差逆伝播法です。2009年には誤差逆伝播法がそれまでの音声認識の水準を上回り、より優れた音声認識能力を備えました。2012年、ディープラーニングを画像認識に適用し、画像認識の技術を競う公式のコンペティションがあり、他の画像認識のテクニックは25%もしくはそれ以上の間違いがありました。私たちの方法でも15%の間違いがありましたが、他に比べると約半分の確率です。2016年、英語・中国語のGoogle翻訳を開発し、ニューラルネットワークは言語の断片を取り込み、別の言語の断片を作成できました。言語的な知識無しに、自動翻訳が行われました。膨大な翻訳の例を見せることによってニューラルネットワークが習得しました。Google翻訳はニューラルネットワークを使い、言語の知識がゼロの状態から誤差逆伝播法のアルゴリズムだけを使い知識を得ます。最近、人間の脳でも同じようなことが行われているのではないかと考える人たちが出てきています。進化的アルゴリズムと異なり、誤差逆伝播法は1つずつ重みの変化を比較する代わりに、重みの変化の影響を算出します。」と答えています。

ヒントン教授は、「AIと人間の未来はどうか？」と問われ、「最も人間らしい部分は、人に同情すること、詩を書くことなど創造的なこと、わずかな意味の違いを理解すること部分を実現するのはまだまだ先の話です。人は創造的な分野と汎用的な知識の面で秀でていると思います。AIにものを認識させ翻訳させることができますが、汎用機能を持たせることは、まだできません。」と答えています。

更に、「汎用型AIは実現するのか？」について、ヒントン教授は、「なるかもしれません。でも、それは遠い将来の話だと思います。汎用知能を備えた正真正銘のAIに到達するには、あと50年はかかるかもしれません。理由はスケールが随分と違うのです。

今のところニューラルネットワークは脳に比べものになりません。Google が作りたいと考えているのは、ウェブ上でドキュメントを検索した場合に、そのドキュメントの内容を理解すると言う技術です。AI をさらに発展させるには、人々はもっと大胆で異なる機能の仕方をする別の種類のニューラルネットワークを研究する必要があると思います。今あるニューラルネットワークが正しいとは限りません。」と将来を展望しました。

・ニューラルネットワークの多層化の発想

機械学習・深層学習の基礎となるニューラルネットワークの多層化の発想は、コンピュータの黎明期からありました。しかし、膨大な計算コストが問題となって、長らくニューラルネットワークというのは低迷していました。

第1回講座でも触れましたけれども、澤井も最初の頃機械翻訳開発にニューラルネットワークを使おうと思いました。が、当時のコンピュータ「IBM 360」にしても性能が遅く、メモリがないということで機械翻訳には到底使えませんでした。

例えば、コンピュータの父の1人と言われる J. V. ノイマンは、著書「人工頭脳と自己増殖」や「自己増殖オートマトンの理論」等の著書で、人間の脳やニューロン（神経細胞）の動きを、自己増殖する自動機械（オートマトン）のプログラムとして実現する方法を記述しています。そして、学習機械の元祖にあたる「パーセプトロン」が登場しました。

・パーセプトロン

1957年米国の心理学者フランク・ローゼンブラット（バッファローのコーネル航空研究所）が「パーセプトロン」を考案しました。ローゼンブラットが設計したパーセプトロンは、重さ数トンの電子機械でしたが、現代テクノロジーの魔法が、それを数行の小さなプログラムに置き換えてしまいました。パーセプトロンは神経細胞を数学的にモデル化したものになります。神経細胞の樹状突起から入ってきた「入力」に、神経細胞体の中で「重み」をかけ、神経細胞の中である一定の数値以上になると「活性化」して「1」という値を、軸索を通して次の神経細胞にシナプス経由で伝えます。神経細胞の中である一定の数値以上にならない場合は全て「0」になってしまいます。神経細胞体の中で、しきいち（閾値）というものを通して、0か1かを次のシナプスに信号が送られていく仕組みです。

・オートエンコーダの深層化

2006年にニューラルネットワークの代表的な研究者であるトロント大学のジェフリー・ヒントンらの研究チームが、制限ボルツマンマシンによるオートエンコーダの深層化(ディープニューラルネットワーク(DNN))に成功し、AIが再び注目を集めるようになりました。これにより、コンピュータが「目」を持つようになり、目を持ったコンピュータが見せる未知の領域が広がり始めました。

同時に、コンピュータのハード性能の急激な進歩や、CPUよりも単純な演算の並列処理に優れた画像処理ユニット(GPU)の低価格化により、2012年頃からは急速に研究が活発となり、第三次人工知能ブームが到来しました。以後GPU等がスーパーコンピュータに搭載され、AI研究開発に不可欠となりました。

米Googleは、囲碁の世界チャンピオン「イ・セドル」を破った「AlphaGo」や、様々なサービスで利用している、Google独自の処理装置「TPU」(Tensor Processing Unit)を発表しました。それによると、TPU(テンスー・プロセッシング・ユニット、Tensor processing unit)はGoogleが開発した機械学習に特化した「特定用途向け集積回路」(ASIC)となっています。

・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機は、ディープラーニング専用チップ「Cloud TPU」を64個を相互接続した「TPUボット」の演算能力は11.5PFLOPS(毎秒1,000兆回の浮動小数点演算が実行できる計算能力)で、計算精度は8ビットの精度で異なるものの、単純な演算回数の比較では、8.16PFLOPSのスーパーコンピュータ「京」を上回っています。

数学の観点から言うと、オートエンコーダを使った学習は出力が入力を再現するような学習です。人間が介在しないので「教師なし学習」と言われています。例えば、3という手書き文字をオートエンコーダの入力層と出力層に同じデータを用いて覚え込ませると、中間(隠れ層)に特徴表現として圧縮されます。

Googleのスーパーコンピュータ8台で動作する機械翻訳機トランスフォーマーという機械翻訳は今まである翻訳よりも優れた性能を出しました。

・量子プロセッサ

2019年3月 Google が量子プロセッサ「Sycamore」を用いて、世界最速のスーパーコンピュータでも1万年かかるとされる処理を、Google の測定結果によると200秒で実行しました。Sycamore は、結構小さなもので絶対温度0度近くの温度で動作するというものになっています。

2. 人間の脳の研究と機械学習

・人間の脳のモデル

人間の脳は100億個以上のニューロンで構成されていると言われています。一説によると800億個だとか、1000億個とかというような話もあります。たくさん神経細胞でできています。

脳のモデルを作る研究には2つ目的があります。1つは脳の仕組みを解明すること、2つ目は優れた情報処理装置を作ることです。優れた情報処理装置を作るとは、脳の中にできる外界の写し「世界像」を自己形成し、思考・行動するような機械を創るというようなことです。

人工知能の立場では優れた情報処理装置を作ることになりますが、AI研究者の多くは「脳の仕組みを解明したい」、ひいては「人間とは何か?」というのを知りたいという人が多いようです。

脳のモデルを作る研究では、脳で神経活動が活発になっている部位の神経代謝や脳血流量の変化を、「fMRI」で測定します。1990年代は脳の10年と呼ばれており、脳科学のいろいろな分野で多大な進歩がみられました。

fMRI (functional Magnetic Resonance Imaging) とは、MRIの技術をもとに、脳を対象とした断層画像を取得し、脳の機能や活動の観察をするための装置・技術です。MRIは放射線を用いず、磁石による磁場と電波を用いて、生体外部から組織を画像化し、おもに全身の組織や臓器の構造の観察する装置・技術です。fMRIは、MRIの一種でありますけれども、例えば言葉を使うとき、脳のどこが働いているのか調べることが可能です。血流がよく流れている脳の活性化部位を「赤」で、あまり血流がない脳の非活性部位を「青」で表示しています。脳の仕組みを解明するという立場から、写真を見ると符号化(エンコード)モデルが働いて、視覚特徴、言語特徴、情動特徴や行動特徴というのを脳活動としてfMRIで画像化できます。動としてfMRIで画像化できます。

優れた情報処理を作るという立場からは、例えば、被験者に「感覚入力の写真を見せ、符号化モデルで脳活動を fMRI で解析し、復号化モデルにかけると「家族、料理、幸せ、楽しそう」というような認識内容が得られます。観測不能の脳の情報処理を、AI 情報を利用して可視化できる訳です。

入力によっては、画像に対してラベル付きのデータを作っていくことが必要になります。AI モデルからラベルなしのデータが大量（プール）に来ると、それに対するラベル付けする画像（クリエ）を選択して、人手でラベルを入力すること（アノテーション）があります。例えば「楽しそう」だとか、「幸せ」そうだというように、画像にラベルをつけてやるわけです。これがラベル付き学習データで、AI モデルはこのラベル付き学習データでモデルの学習を行います。

・脳の視覚野

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ます。脳の視覚野は、視覚情報を処理する大脳皮質の領域です。それは後頭葉にあります。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ます。V1、V2、V4そしてITに至る経路は、形状認識とオブジェクト表現に関連しており、目から入った視覚信号からだんだん輪郭を掴んでいきます。また、長期記憶の保存にも関連していると言われています。

・初期視覚野：V1野

医学的な話ですが、初期視覚野と呼ばれる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す。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りません。この好みのことを選択特徴などと呼びます。V1野細胞の受容野の大きさはおよそ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ます。

・視覚野 V2

二次視覚野(V2)は、前線状皮質とも呼ばれます。視覚野の2番目の主要領域であり、視覚連合野内の最初の領域です。V1からの視覚信号を受信し、V3、V4、およびITに視覚信号を送信します。

・高次視覚野：V4野～IT野

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ます。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ます。V1、V2、V4とだんだん輪郭をつかんでいってその抽象的な概念がつかんでいきます。特徴量予測デコーダ（脳-DNN変換）で予測特徴パターンを得て、脳の信号を深層ニューラルネットワーク（DNN）の視覚特徴量に変換することで、ジェット機とか、カメであるとかが検出できるわけです。

fMRIと機械学習を用いて、脳の活動から心の中でイメージした内容の画像化になっています。これは堀川氏等のATR(株式会社国際電気通信基礎技術研究所)と京都大学の研究者で日本人の方が研究され論文に発表されていました。高次視覚野であるV4野からIT野の細胞では、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ります。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ます。

1980年代にNHK技術研究所の福島邦彦氏が提案した「ネオコグニトロン・アーキテクチャ」で、機能的な事実と生理学的な事実を基にした初期視覚野の細胞群のモデルです。ネオコグニトロン・アーキテクチャは、例えば郵便の手書き文字を認識するに使われました。

・汎用AI

AGI（アーティフィシアル・ジェネラル・インテリジェンス）とは、人間のような汎用能力を持ち多種多様な問題解決ができる汎用AIです。汎用人工知能（AGI）は「強いAI」とも言われ、「弱いAI」と言われる特化型AI（ANI）に対して「人間と同様の感性や思考回路をもつ」人工知能のことを指します。「AI白書」によるとAGIの脳アーキテクチャーはこのような形になっています。ディープラーニングが教師なし学習、そして階層型評価学習、思考・ナビゲーションモジュール、言語理解など全ての機能を統合するモジュールと言います。

・強いAIと弱いAI

強いAIとっていうのはどういう風に違うか見てみましょう。

「強いA I」というのは汎用のA Iで、例えば「鉄腕アトム」のように、ほとんど人間と同じような感性を持って動く人工知能またはロボットです。

「弱いA I」というのは道具としてのA Iで、例えば金融ロボットで、現在稼働しているA Iというのは全てまだ弱いA Iなので道具としてのA Iであると思います。

・機械学習

脳における学習の枠組みに基づくと、機械に学習させる「機械学習」には「教師あり学習」、「教師なし学習」、と「強化学習」の三つの学習の枠組みがあります。脳の部位として、それぞれ小脳、大脳皮質、と大脳基底核と深く関連があると言われていています。

・教師あり学習

「教師あり学習」は、主に、小脳が担う学習機能です。学習者に対し、教示者が明示的に正解を教えたり、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ます。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、適切な出力が出せます。誤差逆伝播法(Back Propagation)が、その代表的手法です。また、正解、若しくは誤りを入力として、未経験入力に対する意志を決定する決定木(DecisionTree)の作成などもあります。

ディープラーニングの基本原理は、損失値をフィードバックし、重みをほんの少し調整することです。誤差逆伝播法は「オプティマイザー」で重みを調整する仕組みです。

・教師なし学習

「教師なし学習」は大脳皮質にある機能です。教師なし学習は、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です。主成分分析などの次元圧縮(Dimensionality Compression)手法が代表例です。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ります。

・強化学習

「強化学習」というのは、基本的には脳の基底核の機能です。

強化学習では、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します。例えば、将棋の場合にこんな手を打ったら最終的に勝った、負けたを教えます。強化学習によって、囲碁とか将棋のA Iが強くなっていきます。

3. 画像認識「目を持ったA I」

画像認識、目を持ったA Iとは、人工の脳神経回路（ニューラルネットワーク）で深層学習するもの例です。例えば、ギリシャのカタコルン港に咲いていた植物が何なのかA Iで画像認識すると、「ブラシの木」で学名は・・・で、分類情報は「界」とか「属」とか、そして最後ブラッシュの木という「種」であるというところまで教えてくれます。「目を持ったA I」は従来の事典に代わり、教科の主体的な深い学びの道具「生き字引」として活躍しています。最近、魚類図鑑や動物図鑑などのA I図鑑「リンネレンズ」がL i n n e(株)からWebで公開されています。

4. 畳み込みニューラルネットワーク(CNN)

2012年9月30日に開催された一般物体認識のコンテスト「I L S V R C」(Image Net Large Scale Visual Recognition Challenge)において、「A l e x N e t」という深い構造を持つC N Nが、従来手法の分類性能を大幅に上回り優勝いたしました。

A l e x N e t は畳み込みニューラル ネットワーク (C N N) の構造の名前です。A l e x N e t はエラー率 15. 3% で優勝し、次点よりも 10. 8% 以上低いものでした。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りました。

5. 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」とは、再帰的な繋りを持つニューラルネットワークです。

「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ます。RNNはどんな長さの文章でも処理できる再帰型ニューラルネットワークです。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ました。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ました。

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りましたが、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ります。機械翻訳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ます。

Googleの自動画像キャプションシステムでは画像から「テーブルの上にコーヒカップがある」という文章(キャプション)に翻訳します。

課題

人間の医師とAI手術ロボット、どちらに命を預ける?について考察し、あなたの考えを800字で説明しなさい。