

特集

1.

深層学習登場前夜とその後

The Night before the Appearance of Deep Learning and Then

佐藤真一

abstract

深層学習がありとあらゆる分野で技術革新を起こしている。深層学習は人工ニューラルネットワークの第三次ブームに相当しているが、これまでのブームとは質的に異なるような広がりを見せている。深層学習はどのように登場したのか、以前のニューラルネットワークとどのような違いがあるのか、その後どのように広がっていったのか、等、現在のカンブリア爆発的な広がりの前までを概観したい。

キーワード：深層学習，ニューラルネットワーク，人工知能，機械学習

1. はじめに

最近、人工知能（AI）という言葉を見かけない日はない。これは、人工知能の能力が飛躍的に高まり、様々な应用到広く使われるようになったからにほかならない。自動運転を支えているのも人工知能だし、チェスはおろか将棋や囲碁でも人間は人工知能に太刀打ちできないし、携帯電話の位置情報で人流を解析するのもAIである。人工知能がブームとなっているのはここ10年の出来事である。この人工知能の爆発的なブームの火付け役となったのが深層学習である。深層学習は現在様々な分野で技術革新を起こしており、今もなお様々な応用が爆発的に広がりつつある状況であり、これらの詳細については本特集の他の記事を御参照頂きたい。本稿では、深層学習がどのように生まれてきたのか、現在の爆発的な広がりの前までを概観したい。

深層学習とは、人工神経回路網（人工ニューラルネットワーク、ニューラルネットワーク、artificial neural network）の一種であり、特に「深い」構造を持つことによりその能力を飛躍的に増大させたニューラルネットワークのことを言う。ニューラルネットワークは、脳の

極めて柔軟な知的処理能力に着目し、脳が膨大な数の神経細胞のネットワークによる情報処理システムとして機能していることから、ソフトウェアなどで人工的に実現した神経細胞（ニューロン）を多数組み合わせさせたネットワークにより、様々な情報処理を行わせようとする計算モデルである。ニューラルネットワークに関する研究には、ニューロンや脳の機能や動作原理を解明しようという生理学的側面、人間の脳により実現されている知能の解明という認知心理学的側面、人間の脳のように柔軟に機能する情報処理機構を構築したいという工学的側面を併せ持つ。深層学習は、特に高度な知的処理を含む情報処理機構の実現という点から広く検討されている。

ニューラルネットワーク研究の歴史は古く、1940年代に遡る。深層学習についての理解を深めるためには、ニューラルネットワーク研究の流れについて振り返ることも重要であろう。そこで本稿では、まずニューラルネットワーク研究の歴史について振り返る。ニューラルネットワーク研究にはこれまで何回かの波があり、現在は第三次ブームに相当する。この波は、実は人工知能研究の波とも相互に関連しており、現在は人工知能研究においても第三次ブームに相当する。加えて、深層学習の出現においても人工知能研究の影響は無視できない。そこで人工知能研究の歴史についても振り返る。これらを踏まえ、深層学習の登場までを振り返りたい。

佐藤真一 正員：フェロー 国立情報学研究所コンテンツ科学研究系
E-mail satoh@nii.ac.jp
Shin'ichi SATOH, Fellow (Digital Content and Media Sciences Research Division,
National Institute of Informatics, Tokyo, 101-8430 Japan).
電子情報通信学会誌 Vol.105 No.5 pp.360-363 2022年5月
©電子情報通信学会 2022

2. ニューラルネットワーク研究の歴史

脳は、膨大な数のニューロンが、ある構造をもって相互に結合し合い、信号を送り合うことにより、全体として高度な情報処理システムとして機能している。1940年代に始まったと考えられるニューラルネットワークの研究では、まずニューロンの動作のモデル化が行われた。実際の生理モデルに従って精密にモデル化を行おうという研究もあるが、McCulloch-Pitts モデル⁽¹⁾は、入力信号の線形和としきい値関数により構成される単純なモデルであり、現在のニューラルネットワーク、ひいては深層学習につながる。Hebb は、ニューロンが興奮した（信号出力を行った）場合に、それに寄与した他のニューロンからの入力とのつながり（シナプス結合強度と呼ばれる）を強めることにより学習が可能であることを示した⁽²⁾。主としてこの二つの研究が現在のニューラルネットワークの基本原則と考えられている。

ニューラルネットワークは、その構造により階層形ネットワークと相互結合形ネットワーク（リカレントネットワーク）とに分けられる。階層形ネットワークは、入力層、複数の中間層と、出力層から成り、ニューロン間の結合は入力層から中間層を介し出力層へと一方向に限られている。フィードフォワードネットワークとも呼ばれる。現在深層学習で検討されているのはほとんどこの構造のものである。一方相互結合形ネットワークはニューロン間の結合にこのような制限がない。現在でもリカレントニューラルネットワーク（RNN）と呼ばれ使われている。階層形ネットワークでも最も単純な構造のものとして、1950年代にRosenblattによりパーセプトロンが提案された⁽³⁾。これは任意の線形関数が表現可能であり、かつデータに基づいて学習することにより必ず正解に収束することが証明され、多くの問題が学習により解決できるのではないかと期待され、第一次のニューラルネットワークブームを巻き起こした。しかしながら、Minskyらにより排他的論理和のような非常に簡単な非線形関数でも学習できないことが指摘され⁽⁴⁾、このブームは終えんを迎える。

1990年頃になると、ニューラルネットワークの第二次ブームが始まる。まずはニューラルネットワークの学習アルゴリズムとしてより高速で正確な学習が可能で一般化デルタルール⁽⁵⁾が提案された。更にこれを階層形ネットワークにおける多数の階層にも適用可能とするため、まず出力層から学習を開始し、順次前段階の層の出力誤りを推定しながら、出力層から入力層に向けて逆向きに、各層で一般化デルタルールに基づく学習を行う誤差逆伝搬アルゴリズム⁽⁶⁾が提案された。これにより、パーセプトロンとは異なり多層の中間層を持つ階層形ネットワークの学習が可能となり、非線形性を持つ複雑な識別面も実現可能となった。深層学習においても、基

本的にこのアルゴリズムが利用されている。ところで、この誤差逆伝搬アルゴリズムの原型である、多層（3層以上）のニューラルネットワークに対する確率的最急降下法（SGD: Stochastic Gradient Descent）による学習アルゴリズムは、1967年に甘利により提案された⁽⁷⁾。IEEE Trans. Electronic Computers 誌に掲載されたものであり、電気による計算機がまだ計算機の一つであった頃である（今の IEEE Trans. Computers）。多層構成を持つ階層形ネットワークは多くの問題に適用可能であることが分かり、かつ識別性能も十分に高かったため、大変広く利用され、ニューラルネットワークを組み込んだ家電製品が多数売り出されるにまで至り、第二次ニューラルネットワークブームの絶頂期を迎える。しかし1990年代中頃には、理論的により強固なSVM等の新たな学習アルゴリズムが提案され、ニューラルネットワークは再び下火となる。

3. 人工知能研究の歴史

人工知能（AI: Artificial Intelligence）という用語が初めて使われたのは、1956年のダートマス会議であるとされている。この会議では、10人の研究者が集い、学習や知能などを計算機で実現する方法について議論された。機械学習や自然言語処理（いずれも用語は異なるが）などと並び、ニューラルネットワークも人工知能に関する話題として議論された。深層学習の大ブレイクの契機となった画像認識については、話題には上がっていなかった。ダートマス会議の参加者でもあったMarvin L. Minsky MIT教授は、1966年のある日、学部学生^(注1)の一人を呼び出し、夏休みの課題に計算機にカメラを接続してそこに写るシーンを説明するプログラムを作れと命じたという⁽⁸⁾（諸説ある）。当時は、普通の人なら誰でもできる画像認識（犬の写真を見て犬だと認識することは誰にでもできる）は学部学生の夏休みの宿題程度の課題であり、人工知能のプロの研究者は他の「知的」な課題、例えばチェスを指すプログラムの作成等に取り組むべきと考えられていた。その後チェスはおろか、将棋や囲碁でも計算機プログラムが人間を負かしてしまったが、現在は画像認識、音声認識、自然言語処理等、普通の人ができることこそが人工知能の重要な課題である。

Logic Theorist（1956年）は、世界初の人工知能プログラムとして知られている。これは数学の自動定理証明システムであり、プリンキピア・マテマティカの定理の一部を自動的に証明できた。ダートマス会議にて公開されたとのことである。また、精神分析医をエミュレートし、人間と対話ができるELIZAも登場した。これらに

(注1) ちなみにこの学部学生はGerald Sussmann、後にMIT教授となるが、画像認識の研究者とはならなかった。

より、人工知能の第一次ブームが起こった。その後、1960年代には、化学分野の知識を搭載し、有機化合物の分析を行う DENDRAL、更には1970年代には医学的知識を搭載し、医師との対話に基づき血液疾患に対する適切な抗生物質を推薦する MYCIN というシステムが提案された。これらは知識をルールとして搭載すれば専門的な問題解決ができるとされたエキスパートシステムである。1980年代に入ると、汎用推論エンジンに専門家の知識をルール化して搭載したエキスパートシステムの有用性に注目が集まり、商用システムまで出現した。これが人工知能の第二次ブームとなった。しかしながら、実用的な大規模システムを構築しようとする、大変な規模のルールを作らなければならない、かつ人間の知識をルールとして外在化させるのもそれほど簡単ではなく、ルールの作成もそのメンテナンスも大変困難であることが分かってきた。これはフレーム問題と呼ばれている。これにより、人工知能の第二次ブームは下火となってしまふ。

4. 深層学習の登場

人工知能の第一次ブームでは、計算機が知的に振る舞うように、いうならば知識をコードに置き換えてプログラミングしようとした。第二次ブームでは、知識をルールとして記述し、それを汎用推論エンジンに搭載することにより人工知能システムを実現しようとした。いずれも、実用的なシステム構築は困難であることが分かり、頓挫してしまう。1990年頃に、知識をコードやルールとして記述する代わりに、事例データのみを準備し、それをニューラルネットワーク、決定木、SVM等の機械学習システムに「食わせ」、暗黙的に知識を獲得するというアプローチが注目を集めるようになる。このようなアプローチは、画像処理分野における手書き文字認識や顔認識をはじめ、音声認識、自然言語処理でも成功を収めていった。

その一方、これらのアプローチの限界も見えてきていた。例えば画像認識においては、手書き文字や顔等の特定の対象については十分な性能を実現できたが、一般の画像の認識はまだ難しかった。こうした研究を推進するために、一般画像認識のためのベンチマークデータが構築された。ImageNet⁽⁹⁾はその一つであり、このデータを用いた ILSVRC (ImageNet Large Scale Visual Recognition Challenge)⁽¹⁰⁾ というコンペでは、100万余りの学習用画像を基に1,000クラスの画像識別器を構築し、その識別性能を測ろうというものであった。2012年までの典型的なアプローチは、各画像から SIFT⁽¹¹⁾等の局所特徴量を大量に求め、それらを統合する VLAD⁽¹²⁾や Fisher vector⁽¹³⁾等のエンコード手法を用いて画像をベクトル化し、その特性に合わせたカーネル関数を設計し、それを用いた SVMにより識別器を学習す

るというような、極めて複雑な方法であった。しかし1,000クラスの識別は困難な問題であり、なかなか識別性能が上がらなかった。そうした中出てきたのが、当時ほとんど注目されていなかったニューラルネットワークに基づく画像認識手法であった⁽¹⁴⁾。2012年のコンペでぶっちぎりの高性能をたたき出したこの方法は、8層という当時では考えられないほど深い階層を持つ階層形ネットワークに基づく方法であるが、ネットワークへの入力には画像をそのまま入ればよく、局所特徴量やエンコーディング等の心配は一切不要であった。これは一大センセーションを巻き起こし、ほどなく画像認識においては局所特徴量+SVMという従来のアプローチは完全に淘汰され、ニューラルネットワーク一色となった。これが深層学習の大ブレークのきっかけの一つとなり、音声認識や自然言語処理等他の分野にも瞬く間に広まっていった。

ちなみに、この方法の重要な要素技術の一つが畳込み層の利用であり、画像認識では特に有効であった。この畳込み構造を持つニューラルネットワークは、実は福島によるネオコグニトロンにより最初に実現されたものである⁽¹⁵⁾。先に紹介した甘利による確率的最急降下法と合わせ、深層学習の重要な要素技術の二つが日本人の手によるものであることは大変興味深い。

5. 深層学習拡大の背景とその問題点

深層学習を成功に導いた一つの大きな要因は、大規模な学習データの存在である。画像処理、音声認識、自然言語処理など、様々な分野で学習データの整備が進められた。これを可能とした要因は幾つも挙げられるが、例えばインターネットの普及、マイクロコンピュータの普及、それに伴う携帯電話やスマートフォンの普及、大容量の記憶装置、カメラをはじめとする様々なセンサの低廉化・普及等が考えられる。様々な要因があいまって、特に2000年以降、大規模なデータ作成はかなり容易になり、これが深層学習の進展を加速したことは間違いない。ただし、大規模データが用意できたとしても、深層学習の学習に利用するためには、教師データの作成が必須である。例えば画像意味分類用のデータであれば、集めた大量の画像に対し、一つ一つに正解の意味ラベルを付与する必要がある。これは通常人間でなければ行うことができず、ここがボトルネックになりかねない。こうした点を解決するための研究、例えば少数の教師付きデータと大量の教師なしデータから効果的な学習を実現する方法の研究等が検討されている。

また、深層学習の実現のためには、一つ一つは単純な動作をする大量のニューロンのネットワークとしての作用のシミュレーションをする必要があり、大規模高並列の計算が必要となる。深層学習としては、数百万から、

億、更には兆に至る数のニューロンを扱う必要がある。深層学習が大ブレイクをしたきっかけとなった今一つの研究として、Google の研究者らによる、1,000 万の YouTube 映像を用い、教師なし学習によりニューラルネットワーク（オートエンコーダ）を学習したところ、猫の顔を学習したという報告がある⁽¹⁶⁾。これには1,000 万台のマシン（1万6,000 コア）を3日間回し続ける必要があったという。このような規模の計算環境をおいそれと使える研究者は僅かであろう。一方、一つ一つのニューロンの動作は単純であり、単純な均一の計算を極めて高並列に実行する SIMD 計算機との親和性が高い。ちょうど2010年頃には、PCによるゲームやグラフィックス用途に、単純な処理を高並列（画素数分）実行するためのGPUが広く使われており、それを通常の計算にも使おうというGPGPU（General Purpose GPU）という技術が広まっていた。GPUを用いることで、一気に数千もの同時並列実行が実現可能となる。そこで、深層学習をGPUで計算しようという動きは一気に広まった。先に紹介した研究⁽¹⁴⁾でもGPUを利用してニューラルネットワークを実現している^(注2)。GPU利用をサポートした深層学習の様々なライブラリも整備され、今やGPUなしで深層学習を実現することはあり得ないほどになっている。

深層学習により今やタスクによっては人間よりも高性能に知的タスクを実行可能となった。その一方、深層学習はエンドツーエンド学習を行っていると言われ、すなわち入力と出力の大量の組から学習するが、その問題解決課程はブラックボックスとなってしまふ。知的問題解決を行うプログラムを人間が作成したり、ルールを記述したりすることを諦め、学習によりシステムを構築しようとしたことが深層学習の成功につながっており、システムの内部がブラックボックスになってしまうのは不可避ではある。しかしそれでは、例えば自動運転のための画像認識を担うシステムにおいて、何らかの問題により事故を起こした場合、その原因が何なのか不明となってしまふ。こうした点を解決するため、深層学習の説明可能性を高める研究が盛んに検討されている。

6. おわりに

深層学習の登場まで、ニューラルネットワーク研究と人工知能研究の歴史との関連性も併せて振り返った。深層学習の重要な構成要素や、深層学習を成功に導いた背景を再認識すると同時に、内在する問題点についても触れた。深層学習は、当初は人間と同様の知的タスクを実

(注2) 利用された GTX 580 という GPU は有するメモリ容量が 3 GByte であり、構築したニューラルネットワーク全体が載らず、半分ずつ動かすように実装したという。

現するために導入されたものであり、画像認識・音声認識・自然言語処理などの知的タスクが極めて高性能に実行可能となった。その一方で、「人間と同様に知的」であるという枠から外れた、画像符号化やレンダリング等への応用も広がっている。今後ますます、より広範囲、より高性能の手法が実現されていくものと期待している。

文 献

- (1) W.S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115-133, 1943.
- (2) D.O. Hebb, *The Organization of Behavior: A Neuropsychological Theory*, Willey & Sons, New York, 1949.
- (3) F. Rosenblatt, "The perceptron: A probabilistic mode for information storage and organization in the brain," *Psychological Review*, vol. 65, no. 6, pp. 386-408, 1958.
- (4) M. Minsky and S. Papert, *Perceptrons; an introduction to computational geometry*, MIT Press, 1969.
- (5) D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing*, D.E. Rumelhart, J.L. McClelland and the PDP Research Group eds., vol. 1, pp. 318-362, The MIT Press, 1986.
- (6) D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533-536, 1986.
- (7) S. Amari, "A theory of adaptive pattern classifiers," *IEEE Trans. Electronic Computers*, vol. EC-16, no. 3, pp. 299-307, 1967.
- (8) R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2010.
- (9) J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," *Proc. Computer Vision and Pattern Recognition*, pp. 248-255, 2009.
- (10) O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211-252, 2015.
- (11) D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91-110, 2004.
- (12) H. Jégou, M. Douze, C. Schmid, and P. Pérez, "Aggregating local descriptors into a compact image representation," *Proc. Computer Vision and Pattern Recognition*, 2010.
- (13) J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: Theory and practice," *Int. J. Comput. Vis.*, vol. 105, no. 3, pp. 222-245, 2013.
- (14) A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.
- (15) 福島邦彦, "位置ずれに影響されないパターン認識機構の神経回路モデル—ネオコグニトロン—," *信学論(A)*, vol. J62-A, no. 10, pp. 658-665, Oct. 1979.
- (16) Q.V. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G.S. Corrado, J. Dean, and A.Y. Ng, "Building high-level features using large scale unsupervised learning," *Proc. International Conference on Machine Learning*, pp. 507-514, 2012.

(2021年12月30日受付)



さとう しんいち
佐藤 真一 (正員:フェロー)

1987 東大・工・電子卒。1992 同大学院工学系研究科情報工学専攻博士課程了。学術情報センター助手等を経て、2004 から国立情報学研究所教授、現在に至る。1995~1997、米国カーネギーメロン大客員研究員として Informedia 映像デジタルライブラリの研究に従事。工博。画像理解、画像データベース、映像データベース等の研究に従事。